



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) Publication number: **0 386 859 B1**

(12)

EUROPEAN PATENT SPECIFICATION

- (42) Date of publication of patent specification: **01.02.95** (51) Int. Cl.⁶: **C12N 15/54, C12N 15/34, C12N 9/12, //C12Q1/68**
- (21) Application number: **90201140.2**
- (22) Date of filing: **24.12.87**
- (60) Publication number of the earlier application in accordance with Art.76 EPC: **0 265 293**

(54) **T7 DNA polymerase.**

- (30) Priority: **14.01.87 US 3227**
14.12.87 US 132569
- (43) Date of publication of application:
12.09.90 Bulletin 90/37
- (45) Publication of the grant of the patent:
01.02.95 Bulletin 95/05
- (54) Designated Contracting States:
AT BE CH DE ES FR GB GR IT LI LU NL SE
- (56) References cited:
- THE JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 262, no. 32, 15th November 1987, pages 15330-15333; S. TABOR et al.: "Selective oxidation of the exonucleasedomain of bacteriophage T7 DNA polymerase"**
- ANALYTICAL CHEMISTRY, vol. 143, 1984, pages 298-303; R.A. McGRAW: "Dideoxy DNA sequencing with end-labeled oligonucleotide primers"**
- PROTEINS: STRUCTURE, FUNCTION, AND GENETICS, vol. 1, no. 1, 1986, pages 66-78; P.S. FREEMONT et al.: "A domain of the klenow**

fragment of Escherichia coli DNA polymerase I has polymerase but no exonuclease activity"

- (73) Proprietor: **THE PRESIDENT AND FELLOWS OF HARVARD COLLEGE**
17 Quincy Street
Cambridge, MA 02138 (US)
- (72) Inventor: **Tabor, Stanley**
9a Lowell Street
Cambridge,
Massachusetts 02138 (US)
Inventor: **Richardson, Charles C.**
78 Chestnut Hill Road
Chestnut Hill,
Massachusetts 02167 (US)
- (74) Representative: **Moon, Donald Keith et al**
BREWER & SON
Quality House
Quality Court
Chancery Lane
London WC2A 1HT (GB)

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid (Art. 99(1) European patent convention).

Description

This invention relates to DNA polymerases suitable for DNA sequencing and in particular relates to a method of producing the purified modified polymerases and to a purified modified gene encoding a modified DNA polymerase.

DNA sequencing involves the generation of four populations of single stranded DNA fragments having one defined terminus and one variable terminus. The variable terminus always terminates at a specific given nucleotide base (either guanine (G), adenine (A), thymine (T), or cytosine (C)). The four different sets of fragments are each separated on the basis of their length, on a high resolution polyacrylamide gel; each band on the gel corresponds colinearly to a specific nucleotide in the DNA sequence, thus identifying the positions in the sequence of the given nucleotide base.

Generally there are two methods of DNA sequencing. One method (Maxam and Gilbert sequencing) involves the chemical degradation of isolated DNA fragments, each labeled with a single radiolabel at its defined terminus, each reaction yielding a limited cleavage specifically at one or more of the four bases (G, A, T or C). The other method (dideoxy sequencing) involves the enzymatic synthesis of a DNA strand. Four separate syntheses are run, each reaction being caused to terminate at a specific base (G, A, T or C) via incorporation of the appropriate chain terminating dideoxynucleotide. The latter method is preferred since the DNA fragments are uniformly labelled (instead of end labelled) and thus the larger DNA fragments contain increasingly more radioactivity. Further, ³⁵S-labelled nucleotides can be used in place of ³²P-labelled nucleotides, resulting in sharper definition; and the reaction products are simple to interpret since each lane corresponds only to either G, A, T or C. The enzyme used for most dideoxy sequencing is the *Escherichia coli* DNA-polymerase I large fragment ("Klenow"). Another polymerase used is AMV reverse transcriptase.

Summary of the Invention

In one aspect the invention features a method for producing a purified modified DNA polymerase which method comprises expressing a modified gene which gene encodes a modified processive DNA polymerase which has sufficient DNA polymerase activity for use in DNA sequencing when said polymerase is combined with any cofactor necessary for said DNA polymerase activity and which results from the modification of a naturally occurring gene modified in that one or more amino acids in the 3' - 5' exonuclease domain of said naturally occurring DNA polymerase are replaced by an amino acid other than that naturally occurring at the site of substitution or are deleted so as to reduce the activity of naturally occurring 3' - 5' exonuclease activity of the naturally occurring DNA polymerase.

In another aspect the invention features a purified modified gene that encodes a processive modified T7-type DNA polymerase which polymerase is able to remain bound to DNA for at least 500 bases under conditions normally used for DNA sequencing reactions and which has sufficient DNA polymerase activity for use in DNA sequencing when said polymerase is combined with any host factor necessary for said DNA polymerase activity and which results from the modification of a naturally occurring gene modified to reduce the activity of naturally occurring 3' - 5' exonuclease activity of the naturally occurring DNA polymerase wherein one or more amino acids of the exonuclease domain within the amino terminal half of the T7 DNA polymerase of said naturally occurring DNA polymerase, or the corresponding domain of other T7-type DNA polymerases, are replaced by an amino acid other than that naturally occurring at the site of substitution or are deleted.

In a preferred embodiment one or more of the amino acids of the exonuclease domain from the terminal to amino acid residue 224 of T7 DNA polymerase of said naturally occurring DNA polymerase the corresponding domain of other T7-type DNA polymerases, are replaced by an amino acid other than that naturally occurring at the site of substitution or are deleted.

In preferred embodiments of the above mentioned aspects of the invention the DNA polymerase encoded by the gene is able to remain bound to a DNA molecule for at least 500 bases in cells normally used for DNA sequencing before dissociating, most preferably for a polymerase is polymerase activity of the modified polymerase is at least 90% of that of a polymerase; the gene encodes a modified polymerase which is substrate specific; the DNA polymerase infected with a T7-type phage (i.e., phage in which the DNA polymerase subunit) for example, the T7-type phage is T7, T3, Φ1, Φ11, H, W31 non-discriminating for dideoxy nucleotide analogs; the polymerase activity of the modified polymerase is 50% lower than the naturally-occurring exonuclease activity of the naturally occurring polymerase; the polymerase is modified to reduce the activity of the

less than 500 units per mg of polymerase, more preferably less than 1 unit, even more preferably less than 0.1 unit, and most preferably has no detectable exonuclease activity; the polymerase is able to utilize primers of as short as 10 bases or preferably as short as 4 bases; the primer comprises four to forty nucleotide bases, and is single stranded DNA or RNA. Preferably the modified gene encodes a modified DNA polymerase in which a naturally occurring His residue of the naturally occurring DNA polymerase is replaced or deleted. In embodiments described herein after His 123, Ser 122 and His 123, Lys 118 to His 123, Lys 118 and Arg 119 or Arg 131, Lys 136, Lys 140, Lys 144 and Arg 145 of the naturally occurring T7 DNA Polymerase are replaced or deleted.

This invention provides a DNA polymerase which is processive, non-discriminating, and can utilize short primers. Further, the polymerase has no associated exonuclease activity. These are ideal properties for the above described methods, and in particular for DNA sequencing reactions, since the background level of radioactivity in the polyacrylamide gels is negligible, there are few or no artifactual bands, and the bands are sharp -- making the DNA sequence easy to read. Further, such a polymerase allows novel methods of sequencing long DNA fragments, as is described in detail below.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof and from the claims.

Description of the Preferred Embodiments

The drawings will first briefly be described.

Drawings

Figs. 1-3 are diagrammatic representations of the vectors pTrx-2, mGP1-1, and pGP5-5 respectively; Fig. 4 is a graphical representation of the selective oxidation of T7 DNA polymerase; Fig. 5 is a graphical representation of the ability of modified T7 polymerase to synthesize DNA in the presence of etheno-dATP; and Fig. 6 is a diagrammatic representation of the enzymatic amplification of genomic DNA using modified T7 DNA polymerase.

Fig. 7, 8 and 9 are the nucleotide sequences of pTrx-2, a part of pGP5-5 and mGP1-2 respectively.

Fig. 10 is a diagrammatic representation of pGP5-6.

DNA Polymerase

In general the DNA polymerase of this invention is processive, has no associated exonuclease activity, does not discriminate against nucleotide analog incorporation, and can utilize small oligonucleotides (such as tetramers, hexamers and octamers) as specific primers. These properties will now be discussed in detail.

Processivity

By processivity is meant that the DNA polymerase is able to continuously incorporate many nucleotides using the same primer-template without dissociating from the template, under conditions normally used for DNA sequencing extension reactions. The degree of processivity varies with different polymerases: some incorporate only a few bases before dissociating (e.g. Klenow (about 15 bases), T4 DNA polymerase (about 10 bases), T5 DNA polymerase (about 180 bases) and reverse transcriptase (about 200 bases) (Dahlmann-Buchdtorf, *Biol. Chem.* 254:1227 1979; Bambara et al., *J. Biol. Chem.* 253:413, 1978) while others, such as the present invention, will remain bound for at least 500 bases and preferably at least 1,000 bases under suitable environmental conditions. Such environmental conditions include having adequate concentrations of the four deoxynucleoside triphosphates and an incubation temperature from 10°C-50°C. The processivity is non-enhanced in the presence of *E. coli* single stranded binding (ssb), protein.

With processive enzymes termination of a sequencing reaction will occur by incubating the reaction mixture with a chain terminating agent, such as a dideoxynucleotide, which "chase" the polymerase, then artifactual bands will arise during sequencing reactions, on where the DNA polymerase dissociates from the template at an incorrect position and obscures the true DNA sequence. This process of "chasing" the polymerase has a high frequency of artifactual bands up to a high molecular weight at the top of the gel, where the DNA sequence is read. This is not an ideal solution since

probability of dissociating from the template at regions of compact secondary structure, or hairpins. Reinitiation of primer elongation at these sites is inefficient and the usual result is the formation of bands at the same position for all four nucleotides, thus obscuring the DNA sequence.

5 Analog discrimination

The DNA polymerases of this invention do not discriminate significantly between dideoxy-nucleotide analogs and normal nucleotides. That is, the chance of incorporation of an analog is approximately the same as that of a normal nucleotide or at least incorporates the analog with at least 1/10 the efficiency that
 10 of a normal analog. The polymerases of this invention also do not discriminate significantly against some other analogs. This is important since, in addition to the four normal deoxynucleoside triphosphates (dGTP, dATP, dTTP and dCTP), sequencing reactions require the incorporation of other types of nucleotide derivatives such as: radioactively-or fluorescently-labelled nucleoside triphosphates, usually for labeling the synthesized strands with ^{35}S , ^{32}P , or other chemical agents. When a DNA polymerase does not discriminate
 15 against analogs the same probability will exist for the incorporation of an analog as for a normal nucleotide. For labelled nucleoside triphosphates this is important in order to efficiently label the synthesized DNA strands using a minimum of radioactivity. Further, lower levels of analogs are required with such enzymes, making the sequencing reaction cheaper than with a discriminating enzyme.

Discriminating polymerases show a different extent of discrimination when they are polymerizing in a processive mode versus when stalled, struggling to synthesize through a secondary structure impediment.
 20 At such impediments there will be a variability in the intensity of different radioactive bands on the gel, which may obscure the sequence.

Exonuclease Activity

25 The DNA polymerase of the invention has less than 50%, preferably less than 1%, and most preferably less than 0.1%, of the normal or naturally associated level of exonuclease activity (amount of activity per polymerase molecule). By normal or naturally associated level is meant the exonuclease activity of unmodified T7-type polymerase. Normally the associated activity is about 5,000 units of exonuclease
 30 activity per mg of polymerase, measured as described below by a modification of the procedure of Chase et al. (249 J. Biol. Chem. 4545, 1974). Exonucleases increase the fidelity of DNA synthesis by excising any newly synthesized bases which are incorrectly basepaired to the template. Such associated exonuclease activities are detrimental to the quality of DNA sequencing reactions. They raise the minimal required concentration of nucleotide precursors which must be added to the reaction since, when the nucleotide
 35 concentration falls, the polymerase activity slows to a rate comparable with the exonuclease activity, resulting in no net DNA synthesis, or even degradation of the synthesized DNA.

More importantly, associated exonuclease activity will cause a DNA polymerase to idle at regions in the template with secondary structure impediments. When a polymerase approaches such a structure its rate of synthesis decreases as it struggles to pass. An associated exonuclease will excise the newly synthesized
 40 DNA when the polymerase stalls. As a consequence numerous cycles of synthesis and excision will occur. This may result in the polymerase eventually synthesizing past the hairpin (with no detriment to the quality of the sequencing reaction); or the polymerase may dissociate from the synthesized strand (resulting in an artifactual band at the same position in all four sequencing reactions); or, a chain terminating agent may be incorporated at a high frequency and produce a wide variability in the intensity of different fragments in a
 45 sequencing gel. This happens because the frequency of incorporation of a chain terminating agent at any given site increases with the number of opportunities the polymerase has to incorporate the chain terminating nucleotide, and so the DNA polymerase will incorporate a chain-terminating agent at a much higher frequency at sites of idling than at other sites.

An ideal sequencing reaction will produce bands of uniform intensity throughout the gel. This is
 50 essential for obtaining the optimal exposure of the X-ray film for every radioactive fragment. If there is variable intensity of radioactive bands, then fainter bands have a chance of going undetected. To obtain uniform radioactive intensity of all fragments, the DNA polymerase should spend the same interval of time at each position on the DNA, showing no preference for either the addition or removal of nucleotides at any given site. This occurs if the DNA polymerase lacks any associated exonuclease, so that it will have only
 55 one opportunity to incorporate a chain terminating nucleotide at each position along the template.

Short primers

The DNA polymerase of the invention is able to utilize primers of 10 bases or less, as well as longer ones, most preferably of 4-20 bases. The ability to utilize short primers offers a number of important advantages to DNA sequencing. The shorter primers are cheaper to buy and easier to synthesize than the usual 15-20-mer primers. They also anneal faster to complementary sites on a DNA template, thus making the sequencing reaction faster. Further, the ability to utilize small (e.g., six or seven base) oligonucleotide primers for DNA sequencing permits strategies not otherwise possible for sequencing long DNA fragments. For example, a kit containing 80 random hexamers could be generated, none of which are complementary to any sites in the cloning vector. Statistically, one of the 80 hexamer sequences will occur an average of every 50 bases along the DNA fragment to be sequenced. The determination of a sequence of 3000 bases would require only five sequencing cycles. First, a "universal" primer (e.g., New England Biolabs #1211, sequence 5' GTAAACGACGGCCAGT 3') would be used to sequence about 600 bases at one end of the insert. Using the results from this sequencing reaction, a new primer would be picked from the kit homologous to a region near the end of the determined sequence. In the second cycle, the sequence of the next 600 bases would be determined using this primer. Repetition of this process five times would determine the complete sequence of the 3000 bases, without necessitating any subcloning, and without the chemical synthesis of any new oligonucleotide primers. The use of such short primers may be enhanced by including gene 2.5 and 4 protein of T7 in the sequencing reaction.

DNA polymerases of this invention, (i.e., having the above properties) include modified T7-type polymerases. That is the DNA polymerase requires host thioredoxin as a sub-unit, and they are substantially identical to a modified T7 DNA polymerase or to equivalent enzymes isolated from related phage, such as T3, ϕ I, ϕ II, H, W31, gh-1, Y, A1122 and SP6. Each of these enzymes can be modified to have properties similar to those of the modified T7 enzyme. It is possible to isolate the enzyme from phage infected cells directly, but preferably the enzyme is isolated from cells which overproduce it. By substantially identical is meant that the enzyme may have amino acid substitutions which do not affect the overall properties of the enzyme. One example of a particularly desirable amino acid substitution is one in which the natural enzyme is modified to remove any exonuclease activity. This modification may be performed at the genetic or chemical level (see below).

Cloning T7 polymerase

As an example of the invention we shall describe the cloning, overproduction, purification, modification and use of T7 DNA polymerase. This processive enzyme consists of two polypeptides tightly complexed in a one to one stoichiometry. One is the phage T7-encoded gene 5 protein of 84,000 daltons (Modrich et al. 150 J. Biol. Chem. 5515, 1975), the other is the *E. coli* encoded thioredoxin, of 12,000 daltons (Tabor et al., J. Biol. Chem. 262:16, 216, 1987). The thioredoxin is an accessory protein and attaches the gene 5 protein (the non-processive actual DNA polymerase) to the primer template. The natural DNA polymerase has a very active 3' to 5' exonuclease associated with it. This activity makes the polymerase useless for DNA sequencing and must be inactivated or modified before the polymerase can be used. This is readily performed, as described below, either chemically, by local oxidation of the exonuclease domain, or genetically, by modifying the coding region of the polymerase gene encoding this activity.

pTrx-2

In order to clone the *trxA* (thioredoxin) gene of *E. coli* wild type *E. coli* DNA was partially cleaved with *Sau*3A and the fragments ligated to *Bam*HI-cleaved T7 DNA isolated from strain T7 ST9 (Tabor et al., in *Thioredoxin and Glutaredoxin Systems: Structure and Function* (Holmgren et al., eds) pp. 285-300, Raven Press, NY; and Tabor et al., *supra*). The ligated DNA was transfected into *E. coli* *trxA*⁻ cells, the mixture plated onto *trxA*⁻ cells, and the resulting T7 plaques picked. Since T7 cannot grow without an active *E. coli* *trxA* gene only those phages containing the *trxA* gene could form plaques. The cloned *trxA* genes were located on a 470 base pair *Hinc*II fragment.

In order to overproduce thioredoxin a plasmid, pTrx-2, was as constructed. Briefly, the 470 base pair *Hinc*II fragment containing the *trxA* gene was isolated by standard procedure (Maniatis et al., *Cloning: A Laboratory Manual*, Cold Spring Harbor Labs., Cold Spring Harbor, N.Y.), and ligated to a derivative of pBR322 containing a *Ptac* promoter (*ptac*-12, Amann et al., 25 Gene 167, 1983). Referring to Fig. 2, *ptac*-12, containing β -lactamase and Col EI origin, was cut with *Pvu*II, to yield a fragment of 2290 bp, which was then ligated to two tandem copies of *trxA* (*Hinc*II fragment) using commercially available linkers (Sma-

BamHI polylinker), to form pTrx-2. The complete nucleotide sequence of pTrx-2 is shown in Figure 7. Thioredoxin production is now under the control of the tac promoter, and thus can be specifically induced, e.g. by IPTG (isopropyl β -D-thiogalactoside).

5 pGP5-5 and mGP1-2

Some gene products of T7 are lethal when expressed in E. coli. An expression system was developed to facilitate cloning and expression of, lethal genes, based on the inducible expression of T7 RNA polymerase. Gene 5 protein is lethal in some E. coli strains and an example of such a system is described
10 by Tabor et al. 82 Proc. Nat. Acad. Sci. 1074 (1985) where T7 gene 5 was placed under the control of the ϕ 10 promoter, and is only expressed when T7 RNA polymerase is present in the cell.

Briefly, pGP5-5 (Fig. 3) was constructed by standard procedures using synthetic BamHI linkers to join T7 fragment from 14306 (NdeI) to 16869 (AhaIII), containing gene 5, to the 560 bp fragment of T7 from 5667 (HincII) to 6166 (Fnu4H1) containing both the ϕ 1.1A and ϕ 1.1B promoters, which are recognized by
15 T7 RNA polymerase, and the 3kb BamHI-HincII fragment of pACYC177 (Chang et al., 134 J. Bacteriol. 1141, 1978). The nucleotide sequence of the T7 inserts and linkers is shown in Fig. 8. In this plasmid gene 5 is only expressed when T7 RNA polymerase is provided in the cell.

Referring to Fig. 3, T7 RNA polymerase is provided on phage vector mGP1-2. This is similar to pGP1-2 (Tabor et al., id.) except that the fragment of T7 from 3133 (HaeIII) to 5840 (HinI), containing T7 RNA
20 polymerase was ligated, using linkers (BglII and SalI respectively), to BamHI-SalI cut M13 mp8, placing the polymerase gene under control of the lac promoter. The complete nucleotide sequence of mGP1-2 is shown in Fig. 9.

Since pGP5-5 and pTrx-2 have different origins of replication (respectively a P15A and a ColE1 origin) they can be transformed into one cell simultaneously. pTrx-2 expresses large quantities of thioredoxin in the presence of IPTG. mGP1-2 can coexist in the same cell as these two plasmids and be used to regulate
25 expression of T7-DNA polymerase from pGP5-5, simply by causing production of T7-RNA polymerase by inducing the lac promoter with, e.g., IPTG.

30 Overproduction of T7 DNA polymerase

There are several potential strategies for overproducing and reconstituting the two gene products of trxA and gene 5. The same cell strains and plasmids can be utilized for all the strategies. In the preferred strategy the two genes are co-overexpressed in the same cell. (This is because gene 5 is susceptible to proteases until thioredoxin is bound to it.) As described in detail below, one procedure is to place the two
35 genes separately on each of two compatible plasmids in the same cell. Alternatively, the two genes could be placed in tandem on the same plasmid. It is important that the T7-gene 5 is placed under the control of a non-leaky inducible promoter, such as ϕ 1.1A, ϕ 1.1B and ϕ 10 of T7, as the synthesis of even small quantities of the two polypeptides together is toxic in most E. coli cells. By non-leaky is meant that less than 500 molecules of the gene product are produced, per cell generation time, from the gene when the
40 promoter, controlling the gene's expression, is not activated. Preferably the T7 RNA polymerase expression system is used although other expression systems which utilize inducible promoters could also be used. A leaky promoter, e.g., plac, allows more than 500 molecules of protein to be synthesized, even when not induced, thus cells containing lethal genes under the control of such a promoter grow poorly and are not suitable in this invention. It is of course possible to produce these products in cells where they are not
45 lethal, for example, the plac promoter is suitable in such cells.

In a second strategy each gene can be cloned and overexpressed separately. Using this strategy, the cells containing the individually overproduced polypeptides are combined prior to preparing the extracts, at which point the two polypeptides form an active T7 DNA polymerase.

50 Example 1: Production of T7 DNA polymerase

E. coli strain 71.18 (Messing et al., Proc. Nat. Acad. Sci. 74:3642, 1977) is used for preparing stocks of mGP1-2. 71.18 is stored in 50% glycerol at -80°C. and is streaked on a standard minimal media agar plate. A single colony is grown overnight in 25 ml standard M9 media at 37°C, and a single plaque of mGP1-2 is
55 obtained by titring the stock using freshly prepared 71.18 cells. The plaque is used to inoculate 10 ml 2X LB (2% Bacto-Tryptone, 1% yeast extract, 0.5% NaCl, 8mM NaOH) containing JM103 grown to an A_{550} = 0.5. This culture will provide the phage stock for preparing a large culture of mGP1-2. After 3-12 hours, the 10 ml culture is centrifuged, and the supernatant used to infect the large (2L) culture. For the

large culture, 4 X 500 ml 2X LB is inoculated with 4 X 5 ml 71.18 cells grown in M9, and is shaken at 37 °C. When the large culture of cells has grown to an $A_{510} = 1.0$ (approximately three hours), they are inoculated with 10 ml of supernatant containing the starter lysate of mGP1-2. The infected cells are then grown overnight at 37 °C. The next day, the cells are removed by centrifugation, and the supernatant is ready to use for induction of K38/pGP5-5/pTrx-2 (see below). The supernatant can be stored at 4 °C for approximately six months, at a titer $\sim 5 \times 10^{11}$ ϕ /ml. At this titer, 1 L of phage will infect 12 liters of cells at an $A_{510} = 5$ with a multiplicity of infection of 15. If the titer is low, the mGP1-2 phage can be concentrated from the supernatant by dissolving NaCl (60 gm/liter) and PEG-6000 (85 gm/liter) in the supernatant, allowing the mixture to settle at 0 °C for 1-72 hours, and then centrifuging (7000 rpm for 20 min). The precipitate, which contains the mGP1-2 phage, is resuspended in approximately 1/20th of the original volume of M9 media.

K38/pGP5-5/pTrx-2 is the *E. coli* strain (genotype HfrC (λ)) containing the two compatible plasmids pGP5-5 and pTrx-2. pGP5-5 plasmid has a P15A origin of replication and expresses the kanamycin (Km) resistance gene. pTrx-2 has a ColEI origin of replication and expresses the ampicillin (Ap) resistance gene. The plasmids are introduced into K38 by standard procedures, selecting Km^R and Ap^R respectively. The cells K38/pGP5-5/pTrx-2 are stored in 50% glycerol at -80 °C. Prior to use they are streaked on a plate containing 50 μ g/ml ampicillin and kanamycin, grown at 37 °C overnight, and a single colony grown in 10 ml LB media containing 50 μ g/ml ampicillin and kanamycin, at 37 °C for 4-6 hours. The 10 ml cell culture is used to inoculate 500 ml of LB media containing 50 μ g/ml ampicillin and kanamycin and shaken at 37 °C overnight. The following day, the 500 ml culture is used to inoculate 12 liters of 2X LB-KPO₄ media (2% Bacto-Tryptone, 1% yeast extract, 0.5% NaCl, 20 mM KPO₄, 0.2% dextrose, and 0.2% casamino acids, pH 7.4), and grown with aeration in a fermentor at 37 °C. When the cells reach an $A_{510} = 5.0$ (i.e. logarithmic or stationary phase cells), they are infected with mGP1-2 at a multiplicity of infection of 10, and IPTG is added (final concentration 0.5 mM). The IPTG induces production of thioredoxin and the T7 RNA polymerase in mGP1-2, and thence induces production of the cloned DNA polymerase. The cells are grown for an additional 2.5 hours with stirring and aeration, and then harvested. The cell pellet is resuspended in 1.5 L 10% sucrose/20 mM Tris-HCl, pH 8.0/25 mM EDTA and re-spun. Finally, the cell pellet is resuspended in 200 ml 10% sucrose/20 mM Tris-HCl, pH 8.0/5 mM EDTA, and frozen in liquid N₂. From 12 liters of induced cells 70 gm of cell paste are obtained containing approximately 700 mg gene 5 protein and 100 mg thioredoxin.

K38/pTrx-2 (K38 containing pTrx-2 alone) overproduces thioredoxin, and it is added as a "booster" to extracts of K38/pGP5-5/pTrx-2 to insure that thioredoxin is in excess over gene 5 protein at the outset of the purification. The K38/pTrx-2 cells are stored in 50% glycerol at -80 °C. Prior to use they are streaked on a plate containing 50 μ g/ml ampicillin, grown at 37 °C for 24 hours, and a single colony grown at 37 °C overnight in 25 ml LB media containing 50 μ g/ml ampicillin. The 25 ml culture is used to inoculate 2 L of 2X LB media and shaken at 37 °C. When the cells reach an $A_{510} = 3.0$, the ptac promoter, and thus thioredoxin production, is induced by the addition of IPTG (final concentration 0.5 mM). The cells are grown with shaking for an additional 12-16 hours at 37 °C, harvested, resuspended in 600 ml 10% sucrose/20 mM Tris-HCl, pH 8.0/25 mM EDTA, and re-spun. Finally, the cells are resuspended in 40 ml 10% sucrose/20 mM Tris-HCl, pH 8.0/5 mM EDTA, and frozen in liquid N₂. From 2L of cells 16 gm of cell paste are obtained containing 150 mg of thioredoxin.

Assays for the polymerase involve the use of single-stranded calf thymus DNA (6mM) as a substrate. This is prepared immediately prior to use by denaturation of double-stranded calf thymus DNA with 50 mM NaOH at 20 °C for 15 min., followed by neutralization with HCl. Any purified DNA can be used as a template for the polymerase assay, although preferably it will have a length greater than 1,000 bases.

The standard T7 DNA polymerase assay used is a modification of the procedure described by Grippo et al. (246 J. Biol. Chem. 6867, 1971). The standard reaction mix (200 μ l final volume) contains 40 mM Tris/HCl pH 7.5, 10 mM MgCl₂, 5 mM dithiothreitol, 100 nmol alkali-denatured calf thymus DNA, 0.3 dGTP, dATP, dCTP and [³H]dTTP (20 cpm/pm), 50 μ g/ml BSA, and varying amounts of T7 DNA polymerase. Incubation is at 37 °C (10 °C-45 °C) for 30 min (5 min-60 min). The reaction is stopped by the addition of 3 ml of cold (0 °C) 1 N HCl-0.1 M pyrophosphate. Acid-insoluble radioactivity is determined by the procedure of Hinkle et al. (250 J. Biol. Chem. 5523, 1974). The DNA is precipitated on ice for 15 min (5 min-12 hr), then precipitated onto glass-fiber filters by filtration. The filters are washed five times with 4 ml of cold (0 °C) 0.1M HCl-0.1M pyrophosphate, and twice with cold (0 °C) 90% ethanol. After drying, the radioactivity on the filters is counted using a non-aqueous scintillation fluor.

One unit of polymerase activity catalyzes the incorporation of 10 nmol of total nucleotide into an acid-soluble form in 30 min at 37 °C, under the conditions given above. Native T7 DNA polymerase and modified T7 DNA polymerase (see below) have the same specific polymerase activity \pm 20%, which ranges between 5,000-20,000 units/mg for native and 5,000-50,000 units/mg for modified polymerase) depending upon the

preparation, using the standard assay conditions stated above.

T7 DNA polymerase is purified from the above extracts by precipitation and chromatography techniques. An example of such a purification follows.

An extract of frozen cells (200 ml K38/pGP5-5/pTrx-2 and 40 ml K38/pTrx-2) are thawed at 0°C overnight. The cells are combined, and 5 ml of lysozyme (15 mg/ml) and 10 ml of NaCl (5M) are added. After 45 min at 0°C, the cells are placed in a 37°C water bath until their temperature reaches 20°C. The cells are then frozen in liquid N₂. An additional 50 ml of NaCl (5M) is added, and the cells are thawed in a 37°C water bath. After thawing, the cells are gently mixed at 0°C for 80 min. The lysate is centrifuged for one hr at 35,000 rpm in a Beckman 45Ti rotor. The supernatant (250 ml) is fraction I. It contains approximately 700 mg gene 5 protein and 250 mg of thioredoxin (a 2:1 ratio thioredoxin to gene 5 protein).

90 gm of ammonium sulphate is dissolved in fraction I (250 ml) and stirred for 60 min. The suspension is allowed to sit for 60 min, and the resulting precipitate collected by centrifugation at 8000 rpm for 60 min. The precipitate is redissolved in 300 ml of 20 mM Tris-HCl pH 7.5/5 mM 2-mercaptoethanol/0.1 mM EDTA/10% glycerol (Buffer A). This is fraction II.

A column of Whatman DE52 DEAE (12.6 cm² x 18 cm) is prepared and washed with Buffer A. Fraction II is dialyzed overnight against two changes of 1 L of Buffer A each until the conductivity of Fraction II has a conductivity equal to that of Buffer A containing 100 mM NaCl. Dialyzed Fraction II is applied to the column at a flow rate of 100 ml/hr, and washed with 400 ml of Buffer A containing 100 NaCl. Proteins are eluted with a 3.5 L gradient from 100 to 400 mM NaCl in Buffer A at a flow rate of 60 ml/hr. Fractions containing T7 DNA polymerase, which elutes at 200 mM NaCl, are pooled. This is fraction III (190 ml).

A column of Whatman P11 phosphocellulose (12.6 cm² x 11 cm) is prepared and washed with 20 mM KPO₄ pH 7.4/5 mM 2-mercaptoethanol/0.1 mM EDTA/10 % glycerol (Buffer B). Fraction III is diluted 2-fold (380 ml) with Buffer B, then applied to the column at a flow rate of 60 ml/hr and washed with 200 ml of Buffer B containing 100mM KCl. Proteins are eluted with a 1.8 L gradient from 100 to 400 mM KCl in Buffer B at a flow rate of 60 ml/hr. Fractions containing T7 DNA polymerase which elutes at 300 KCl, are pooled. This is fraction IV (370 ml).

A column of DEAE-Sephadex A-50 (4.9 cm² x 15 cm) is prepared and washed with 20 mM Tris-HCl 7.0/0.1 mM dithiothreitol/0.1 mM EDTA/10% glycerol (Buffer C). Fraction IV is dialyzed against two changes of 1 L Buffer C to a final conductivity equal to that of Buffer C containing 100 mM NaCl. Dialyzed fraction IV is applied to the column at a flow rate of 40 ml/hr, and washed with 150 ml of Buffer C containing 100 mM NaCl. Proteins are eluted with a 1 L gradient from 100 to 300 mM NaCl in Buffer C at a flow rate of 40 ml/hr. Fractions containing T7 DNA polymerase, which elutes at 210 mM NaCl, are pooled. This is fraction V (120 ml).

A column of BioRad HTP hydroxylapatite (4.9 cm² x 15 cm) is prepared and washed with 20 mM KPO₄, pH 7.4/10 mM 2-mercaptoethanol/2 mM Na citrate/10% glycerol (Buffer D). Fraction V is dialyzed against two changes of 500 ml Buffer D each. Dialyzed fraction V is applied to the column at a flow rate of 30 ml/hr, and washed with 100 ml of Buffer D. Proteins are eluted with a 900 ml gradient from 0 to 180 mM KPO₄, pH 7.4 in Buffer D at a flow rate of 30 ml/hr. Fractions containing T7 DNA polymerases which elutes at 50 mM KPO₄, are pooled. This is fraction VI (130 ml). It contains 270 mg of homogeneous T7 DNA polymerase.

Fraction VI is dialyzed versus 20 mM KPO₄ pH 7.4/0.1 mM dithiothreitol/0.1 mM EDTA/50% glycerol. This is concentrated fraction VI (~65 ml, 4 mg/ml), and is stored at -20°C.

The isolated T7 polymerase has exonuclease activity associated with it. As stated above this must be inactivated. An example of inactivation by chemical modification follows.

Concentrated fraction VI is dialyzed overnight against 20 mM KPO₄ pH 7.4/0.1 mM dithiothreitol/10% glycerol to remove the EDTA present in the storage buffer. After dialysis, the concentration is adjusted to 2 mg/ml with 20 mM KPO₄ pH 7.4/0.1 mM dithiothreitol/10% glycerol, and 30 ml (2mg/ml) aliquots are placed in 50 ml polypropylene tubes. (At 2 mg/ml, the molar concentration of T7 DNA polymerase is 22 μM.)

Dithiothreitol (DTT) and ferrous ammonium sulfate (Fe(NH₄)₂(SO₄)₂·6H₂O) are prepared fresh immediately before use, and added to a 30 ml aliquot of T7 DNA polymerase, to concentrations of 5 mM DTT (0.6 ml of a 250 stock) and 20μM Fe(NH₄)₂(SO₄)₂·6H₂O (0.6 ml of a 1 mM stock). During modification the molar concentrations of T7 DNA polymerase and iron are each approximately 20 μM, while DTT is in 250X molar excess.

The modification is carried out at 0°C under a saturated oxygen atmosphere as follows. The reaction mixture is placed on ice within a dessicator, the dessicator is purged of air by evacuation and subsequently filled with 100% oxygen. This cycle is repeated three times. The reaction can be performed in air (20% oxygen), but occurs at one third the rate.

The time course of loss of exonuclease activity is shown in Fig. 4. ^3H -labeled double-stranded DNA (6 cpm/pmol) was prepared from bacteriophage T7 as described by Richardson (15 J. Molec. Biol. 49, 1966). ^3H -labeled single-stranded T7 DNA was prepared immediately prior to use by denaturation of double-stranded ^3H -labeled T7 DNA with 50 mM NaOH at 20 °C for 15 min, followed by neutralization with HCl.

5 The standard exonuclease assay used is a modification of the procedure described by Chase et al. (*supra*). The standard reaction mixture (100 μl final volume) contained 40 mM Tris/HCl pH 7.5, 10 mM MgCl_2 , 10 mM dithiothreitol, 60 nmol ^3H -labeled single-stranded T7 DNA (6 cpm/pmol), and varying amounts of T7 DNA polymerase. ^3H -labeled double-stranded T7 DNA can also be used as a substrate. Also, any uniformly radioactively labeled DNA, single- or double-stranded, can be used for the assay. Also, 3' end labeled

10 single- or double-stranded DNA can be used for the assay. After incubation at 37 °C for 15 min, the reaction is stopped by the addition of 30 μl of BSA (10mg/ml) and 25 μl of TCA (100% w/v). The assay can be run at 10 °C-45 °C for 1-60 min. The DNA is precipitated on ice for 15 min (1 min - 12 hr), then centrifuged at 12,000 g for 30 min (5 min - 3 hr). 100 μl of the supernatant is used to determine the acid-soluble radioactivity by adding it to 400 μl water and 5 ml of aqueous scintillation cocktail.

15 One unit of exonuclease activity catalyzes the acid solubilization of 10 nmol of total nucleotide in 30 min under the conditions of the assay. Native T7 DNA polymerase has a specific exonuclease activity of 5000 units/mg, using the standard assay conditions stated above. The specific exonuclease activity of the modified T7 DNA polymerase depends upon the extent of chemical modification, but ideally is at least 10-100-fold lower than that of native T7 DNA polymerase, or 500 to 50 or less units/mg using the standard

20 assay conditions stated above. When double stranded substrate is used the exonuclease activity is about 7-fold higher.

Under the conditions outlined, the exonuclease activity decays exponentially, with a half-life of decay of eight hours. Once per day the reaction vessel is mixed to distribute the soluble oxygen, otherwise the reaction will proceed more rapidly at the surface where the concentration of oxygen is higher. Once per day

25 2.5 mM DTT (0.3 ml of a fresh 250 mM stock to a 30 ml reaction) is added to replenish the oxidized DTT.

After eight hours, the exonuclease activity of T7 DNA polymerase has been reduced 50%, with negligible loss of polymerase activity. The 50% loss may be the result of the complete inactivation of exonuclease activity of half the polymerase molecules, rather than a general reduction of the rate of exonuclease activity in all the molecules. Thus, after an eight hour reaction all the molecules have normal

30 polymerase activity, half the molecules have normal exonuclease activity, while the other half have <0.1% of their original exonuclease activity.

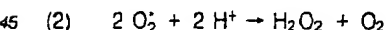
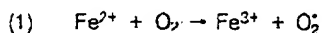
When 50% of the molecules are modified (an eight hour reaction), the enzyme is suitable, although suboptimal, for DNA sequencing. For more optimum quality of DNA sequencing, the reaction is allowed to proceed to greater than 99% modification (having less than 50 units of exonuclease activity), which requires

35 four days.

After four days, the reaction mixture is dialyzed against 2 changes of 250 ml of 20 mM KPO_4 pH 7.4/0.1 mM dithiothreitol/0.1 mM EDTA/50% glycerol to remove the iron. The modified T7 DNA polymerase (~4 mg/ml) is stored at -20 °C.

The reaction mechanism for chemical modification of T7 DNA polymerase depends upon reactive

40 oxygen species generated by the presence of reduced transition metals such as Fe^{2+} and oxygen. A possible reaction mechanism for the generation of hydroxyl radicals is outlined below:



In equation 1, oxidation of the reduced metal ion yields superoxide radical, $\text{O}_2^{\cdot -}$. The superoxide radical

50 can undergo a dismutation reaction, producing hydrogen peroxide (equation 2). Finally, hydrogen peroxide can react with reduced metal ions to form hydroxyl radicals, OH^{\cdot} (the Fenton reaction, equation 3). The oxidized metal ion is recycled to the reduced form by reducing agents such as dithiothreitol (DTT).

These reactive oxygen species probably inactivate proteins by irreversibly chemically altering specific amino acid residues. Such damage is observed in SDS-PAGE of fragments of gene 5 produced by CNBr or

55 trypsin. Some fragments disappear, high molecular weight cross linking occurs, and some fragments are broken into two smaller fragments.

As previously mentioned, oxygen, a reducing agent (e.g. DTT, 2-mercaptoethanol) and a transition metal (e.g. iron) are essential elements of the modification reaction. The reaction occurs in air, but is

stimulated three-fold by use of 100% oxygen. The reaction will occur slowly in the absence of added transition metals due to the presence of trace quantities of transition metals (1-2 μ M) in most buffer preparations.

As expected, inhibitors of the modification reaction include anaerobic conditions (e.g., N₂) and metal chelators (e.g. EDTA, citrate, nitrilotriacetate). In addition, the enzymes catalase and superoxide dismutase may inhibit the reaction, consistent with the essential role of reactive oxygen species in the generation of modified T7 DNA polymerase.

As an alternative procedure, it is possible to genetically mutate the T7 gene 5 to specifically inactivate the exonuclease domain of the protein. The T7 gene 5 protein purified from such mutants is ideal for use in DNA sequencing without the need to chemically inactivate the exonuclease by oxidation and without the secondary damage that inevitably occurs to the protein during chemical modification.

Genetically modified T7 DNA polymerase can be isolated by randomly mutagenizing the gene 5 and then screening for those mutants that have lost exonuclease activity, without loss of polymerase activity. Mutagenesis is performed as follows. Single-stranded DNA containing gene 5 (e.g., cloned in pEMBL-8, a plasmid containing an origin for single stranded DNA replication) under the control of a T7 RNA polymerase promoter is prepared by standard procedure, and treated with two different chemical mutagens: hydrazine which will mutate C's and T's, and formic acid, which will mutate G's and A's. Myers et al. 229 Science 242, 1985. The DNA is mutagenized at a dose which results in an average of one base being altered per plasmid molecule. The single-stranded mutagenized plasmids are then primed with a universal 17-mer primer (see above), and used as templates to synthesize the opposite strands. The synthesized strands contain randomly incorporated bases at positions corresponding to the mutated bases in the templates. The double-stranded mutagenized DNA is then used to transform the strain K38/pGP1-2, which is strain K38 containing the plasmid pGP1-2 (Tabor et al., *supra*). Upon heat induction this strain expresses T7 RNA polymerase. The transformed cells are plated at 30 °C, with approximately 200 colonies per plate.

Screening for cells having T7 DNA polymerase lacking exonuclease activity is based upon the following finding. The 3' to 5' exonuclease of DNA polymerases serves a proofreading function. When bases are misincorporated, the exonuclease will remove the newly incorporated base which is recognized as "abnormal". This is the case for the analog of dATP, etheno-dATP, which is readily incorporated by T7 DNA polymerase in place of dATP. However, in the presence of the 3' to 5' exonuclease of T7 DNA polymerase, it is excised as rapidly as it is incorporated, resulting in no net DNA synthesis. As shown in figure 6, using the alternating copolymer poly d(AT) as a template, native T7 DNA polymerase catalyzes extensive DNA synthesis only in the presence of dATP, and not etheno-dATP. In contrast, modified T7 DNA polymerase, because of its lack of an associated exonuclease, stably incorporates etheno-dATP into DNA at a rate comparable to dATP. Thus, using poly d(AT) as a template, and dTTP and etheno-dATP as precursors, native T7 DNA polymerase is unable to synthesize DNA from this template, while T7 DNA polymerase which has lost its exonuclease activity will be able to use this template to synthesize DNA.

The procedure for lysing and screening large number of colonies is described in Raetz (72 Proc. Nat. Acad. Sci. 2274, 1975). Briefly, the K38/pGP1-2 cells transformed with the mutagenized gene 5-containing plasmids are transferred from the petri dish, where they are present at approximately 200 colonies per plate, to a piece of filter paper ("replica plating"). The filter paper discs are then placed at 42 °C for 60 min to induce the T7 RNA polymerase, which in turn expresses the gene 5 protein. Thioredoxin is constitutively produced from the chromosomal gene. Lysozyme is added to the filter paper to lyse the cells. After a freeze thaw step to ensure cell lysis, the filter paper discs are incubated with poly d(AT), [α -³²P]dTTP and etheno-dATP at 37 °C for 60 min. The filter paper discs are then washed with acid to remove the unincorporated [³²P]dATP. DNA will precipitate on the filter paper in acid, while nucleotides will be soluble. The washed filter paper is then used to expose X-ray film. Colonies which have induced an active T7 DNA polymerase which is deficient in its exonuclease will have incorporated acid-insoluble ³²P, and will be visible by autoradiography. Colonies expressing native T7 DNA polymerase, or expressing a T7 DNA polymerase defective in polymerase activity, will not appear on the autoradiograph.

Colonies which appear positive are recovered from the master petri dish containing the original colonies. Cells containing each potential positive clone will be induced on a larger scale (one liter) and T7 DNA polymerase purified from each preparation to ascertain the levels of exonuclease associated with each mutant. Those low in exonuclease are appropriate for DNA sequencing.

Directed mutagenesis may also be used to isolate genetic mutants in the exonuclease domain of the T7 gene 5 protein. The following is an example of this procedure.

T7 DNA polymerase with reduced exonuclease activity (modified T7 DNA polymerase) can also be distinguished from native T7 DNA polymerase by its ability to synthesize through regions of secondary structure. Thus, with modified DNA polymerase, DNA synthesis from a labeled primer on a template having

secondary structure will result in significantly longer extensions, compared to unmodified or native DNA polymerase. This assay provides a basis for screening for the conversion of small percentages of DNA polymerase molecules to a modified form.

The above assay was used to screen for altered T7 DNA polymerase after treatment with a number of chemical reagents. Three reactions resulted in conversion of the enzyme to a modified form. The first is treatment with iron and a reducing agent, as described above. The other two involve treatment of the enzyme with photooxidizing dyes, Rose Bengal and methylene blue, in the presence of light. The dyes must be titrated carefully, and even under optimum conditions the specificity of inactivation of exonuclease activity over polymerase activity is low, compared to the high specificity of the iron-induced oxidation. Since these dyes are quite specific for modification of histidine residues, this result strongly implicates histidine residues as an essential species in the exonuclease active site.

There are 23 histidine residues in T7 gene 5 protein. Eight of these residues lie in the amino half of the protein, in the region where, based on the homology with the large fragment of *E. coli* DNA polymerase I, the exonuclease domain may be located (Ollis et al. Nature 313, 818, 1984). As described below, seven of the eight histidine residues were mutated individually by synthesis of appropriate oligonucleotides, which were then incorporated into gene 5. These correspond to mutants 1, and 6-10 in table 1.

The mutations were constructed by first cloning the T7 gene 5 from pGP5-3 (Tabor et al., J. Biol. Chem. 262, 1987) into the SmaI and HindIII sites of the vector M13 mp18, to give mGP5-2. (The vector used and the source of gene 5 are not critical in this procedure.) Single-stranded mGP5-2 DNA was prepared from a strain that incorporates deoxyuracil in place of deoxythymidine (Kunkel, Proc. Natl. Acad. Sci. USA 82, 488, 1985). This procedure provides a strong selection for survival of only the synthesized strand (that containing the mutation) when transfected into wild-type *E. coli*, since the strand containing uracil will be preferentially degraded.

Mutant oligonucleotides, 15-20 bases in length, were synthesized by standard procedures. Each oligonucleotide was annealed to the template extended using native T7 DNA polymerase and ligated using T4 DNA ligase. Covalently closed circular molecules were isolated by agarose gel electrophoresis run in the presence of 0.5 µg/ml ethidium bromide. The resulting purified molecules were then used to transform *E. coli* 71.18. DNA from the resulting plaques was isolated and the relevant region sequenced to confirm each mutation.

The following summarizes the oligonucleotides used to generate genetic mutants in the gene 5 exonuclease. The mutations created are underlined. Amino acid and base pair numbers are taken from Dunn et al., 166 J. Molec. Biol. 477, 1983. The relevant wild type sequences of the region of gene 5 mutated are also shown.

35

40

45

50

55

Wild type sequenc :

109 (aa) 122 123
 5 Leu Leu Arg Ser Gly Lys Leu Pro Gly Lys Arg Phe Gly Ser His Ala Leu Glu
 CTT CTG CGT TCC GGC AAG TTG CCC GGA AAA CGC TTT GGG TCT CAC GCT TTG GAG
 14677 (17 bp)

10

Mutation 1: His 123 → Ser 123

15 Primer used: 5' CGC TTT GGA TCC TCC GCT TTG 3'

Mutant sequence:

123
 20 Leu Leu Arg Ser Gly Lys Leu Pro Gly Lys Arg Phe Gly Ser ~~Ser~~ Ala Leu Glu
 CTT CTG CGT TCC GGC AAG TTG CCC GGA AAA CGC TTT GGA TCC TCC GCT TTG GAG

20

Mutation 2: Deletion of Ser 122 and His 123

25 Primer used: 5' GGA AAA CGC TTT GGC GCG TTG GAG GCG 3'

Δ
 6 base deletion

25

Mutant sequence:

122 123
 30 Leu Leu Arg Ser Gly Lys Leu Pro Gly Lys Arg Phe Gly Ala Leu Glu
 CTT CTG CGT TCC GGC AAG TTG CCC GGA AAA CGC TTT GGC --- --- GCG TTG GAG

30

35

40

45

50

55

Mutation 3: Ser 122, His 123 → Ala 122, Glu 123

Primer used: 5' CGC TTT GGG GCT GAG GCT TTG G 3'

Mutant sequence:

Leu Leu Arg Ser Gly Lys Leu Pro Gly Lys Arg Phe Gly ^{122 123} Ala Glu Ala Leu Glu
CTT CTG CGT TCC GGC AAG TTG CCC GGA AAA CGC TTT GGG GCT GAG GCT TTG GAG

Mutation 4: Lys 118, Arg 119 → Glu 118, Glu 119

Primer used: 5' 5' G CCC GGG GAA GAG TTT GGG TCT CAC GC 3'

Mutant sequence:

Leu Leu Arg Ser Gly Lys Leu Pro Gly ^{118 119} Glu Glu Phe Gly Ser His Ala Leu Glu
CTT CTG CGT TCC GGC AAG TTG CCC GGG GAA GAG TTT GGG TCT CAC GCT TTG GAG

Mutation 5: Arg 111, Ser 112, Lys 114 → Glu 111, Ala 112, Glu 114

Primer used: 5' G GGT CTT CTG GAA GCC GGC GAG TTG CCC GG 3'

Mutant sequence:

Leu Leu ^{111 112 114} Glu Ala Gly Glu Leu Pro Gly Lys Arg Phe Gly Ser His Ala Leu
Glu
CTT CTG GAA GCC GGC GAG TTG CCC GGA AAA CGC TTT GGG TCT CAC GCT TTG GAG

Mutation 6: His 59, His 62 → Ser 59, Ser 62

Primer used: 5' ATT GTG TTC TCC AAC GGA TCC AAG TAT GAC G 3'

Wild-type sequence:

aa: 55 59 62
Leu Ile Val Phe His Asn Gly His Lys Tyr Asp Val
CTT ATT GTG TTC CAC AAC GGT CAC AAG TAT GAC GTT
T7 bp: 14515

Mutant sequence:

Leu Ile Val Phe ^{59 62} Ser Asn Gly Ser Lys Tyr Asp Val
CTT ATT GTG TTC TCC AAC GGA TCC AAG TAT GAC GTT

Mutation 7: His 82 → Ser 82

Primer used: 5' GAG TTC ~~TCC~~ CTT CCT CG 3'

Wild-type sequence:

aa: 77 82
 Leu Asn Arg Glu Phe His Leu Pro Arg Glu Asn
 TTG AAC CGA GAG TTC CAC CTT CCT CGT GAG AAC
 T7 bp: 14581

Mutant sequence:

82
 Leu Asn Arg Glu Phe ~~Ser~~ Leu Pro Arg Glu Asn
 TTG AAC CGA GAG TTC ~~TCC~~ TTT CCT CGT GAG AAC

Mutation 8: Arg 96, His 99 → Leu 96, Ser 99

Primer used: 5' CTG TTG ATT TCT TCC AAC CTC 3'

Wild-type sequence:

aa: 93 96 99
 Val Leu Ser Arg Leu Ile His Ser Asn Leu Lys Asp Thr Asp
 GTG TTG TCA CGT TTG ATT CAT TCC AAC CTC AAG GAC ACC GAT
 T7 bp: 14529

Mutant sequence:

96 99
 Val Leu Ser ~~Leu~~ Leu Ile ~~Ser~~ Ser Asn Leu Lys Asp Thr Asp
 GTG TTG TCA CTG TTG ATT TCT TCC AAC CTC AAG GAC ACC GAT

Mutation 9: His 190 → Ser 190

Primer used: 5' CT GAC AAA TCT TAC TTC CCT 3'

Wild-type sequence:

aa: 185 190
 Leu Leu Ser Asp Lys His Tyr Phe Pro Pro Glu
 CTA CTC TCT GAC AAA CAT TAC TTC CCT CCT GAG
 T7 bp: 14905

Mutant sequence:

190
 Leu Leu Ser Asp Lys ~~Ser~~ Tyr Phe Pro Pro Glu
 CTA CTC TCT GAC AAA TCT TAC TTC CCT CCT GAG

Mutation 10: His 218 → Ser 218

Primer used: 5' GAC ATT GAA TCT CGT GCT GC 3'

Wild-type sequence:

aa: 214 218
Val Asp Ile Glu His Arg Ala Ala Trp Leu Leu
GTT GAC ATT GAA CAT CGT GCT GCA TGG CTG CTC
T7 bp: 14992

Mutant sequence:

218
Val Asp Ile Glu ~~Ser~~ Arg Ala Ala Trp Leu Leu
GTT GAC ATT GAA ~~TCT~~ CGT GCT GCA TGG CTG CTC

Mutation 11: Deletion of amino acids 118 to 123

Primer used: 5' C GGC AAG TTG CCC GGG GCT TTG GAG GCG TGG G 3'

Δ
18 base deletion

Wild-type sequence:

109 (aa) 118 122 123 126
Leu Leu Arg Ser Gly Lys Leu Pro Gly Lys Arg Phe Gly Ser His Ala Leu Glu
CTT CTG CGT TCC GGC AAG TTG CCC GGA AAA CGC TTT GGG TCT CAC GCT TTG GAG
14677 (T7 bp)

Mutant sequence:

117 124
Leu Leu Arg Ser Gly Lys Leu Pro Gly.....(6 amino acids).....Ala Leu Glu
CTT CTG CGT TCC GGC AAG TTG CCC GGG.....(18 bases).....GCT TTG GAG

Mutation 12: His 123 → Glu 123

Primer used: 5' GGG TCT GAG GCT TTG G 3'

Mutant sequence:

123
Leu Leu Arg Ser Gly Lys Leu Pro Gly Lys Arg Phe Gly Ser ~~Glu~~ Ala Leu Glu
CTT CTG CGT TCC GGC AAG TTG CCC GGA AAA CGC TTT GGG TCT GAG GCT TTG GAG

Mutation 13: (Arg 131, Lys 136, Lys 140, Lys 144, Arg 145 →
Glu 131, Glu 136, Glu 140, Glu 144, Glu 145)

Primer used: 5' GGT TAT GAG ~~CTC~~ GGC GAG ATG GAG GGT GAA TAC GAA GAC GAC TTT GAG GAA ATG
CTT GAA G 3'

Wild-type sequence:

129(aa) 131 136 140 144 145
Gly Tyr Arg Leu Gly Glu Met Lys Gly Glu Tyr Lys Asp Asp Phe Lys Arg Met Leu Glu Glu
GGT TAT CGC TTA GGC GAG ATG AAG GGT GAA TAC AAA GAC GAC TTT AAG CGT ATG CTT GAA G
14737 (T7 bp)

Mutant sequence:

129(aa) 131 136 140 144 145
Gly Tyr Glu Leu Gly Glu Met Glu Gly Glu Tyr Glu Asp Asp Phe Glu Glu Met Leu Glu Glu
GGT TAT GAG CTC GGC GAG ATG GAG GGT GAA TAC GAA GAC GAC TTT GAG GAA ATG CTT GAA G
14737 (T7 bp)

Each mutant gene 5 protein was produced by infection of the mutant phage into K38/pGP1-2, as follows. The cells were grown at 30°C to an $A_{590} = 1.0$. The temperature was shifted to 42°C for 30 min., to induce T7 RNA polymerase. IPTG was added to 0.5 mM, and a lysate of each phage was added at a moi=10. Infected cells were grown at 37°C for 90 min. The cells were then harvested and extracts prepared by standard procedures for T7 gene 5 protein.

Extracts were partially purified by passage over a phosphocellulose and DEAE A-50 column, and assayed by measuring the polymerase and exonuclease activities directly, as described above. The results are shown in Table 1.

Table 1
SUMMARY OF EXONUCLEASE AND POLYMERASE
ACTIVITIES OF T7 GENE 5 MUTANTS

Mutant	Exonuclease activity, %	Polymerase activity, %
[Wild-type]	[100] ^a	[100] ^b
Mutant 1 (His 123 → Ser 123)	10-25	>90
Mutant 2 (Δ Ser 122, His 123)	0.2-0.4	>90
Mutant 3 (Ser 122, His 123 → Ala 122, Glu 123)	<2	>90

Table 1
SUMMARY OF EXONUCLEASE AND POLYMERASE
ACTIVITIES OF T7 GENE 5 MUTANTS

	Mutant	Exonuclease activity, %	Polymerase activity, %
5	Mutant 4		
10	(Lys 118, Arg 119 → Glu 118, Glu 119)	<30	>90
	Mutant 5		
	(Arg 111, Ser 112, Lys 114 → Glu 111, Ala 112, Glu 114)	>75	>90
15	Mutant 6		
	(His 59, His 62 → Ser 59, Ser 62)	>75	>90
	Mutant 7		
20	(His 82 → Ser 82)	>75	>90
	Mutant 8		
	(Arg 96, His 99 → Leu 96, Ser 99)	>75	>90
	Mutant 9		
25	(His 190 → Ser 190)	>75	>90
	Mutant 10		
	(His 218 → Ser 218)	>75	>90
30	Mutant 11		
	(Δ Lys 118, Arg 119, Phe 120, Gly 121, Ser 122, His 123)	<0.02	>90
	Mutant 12		
35	(His 123 → Glu 123)	<30	>90
	Mutant 13		
40	(Arg 131, Lys 136, Lys 140, Lys 144, Arg 145 → Glu 131, Glu 136, Glu 140, Glu 144, Glu 145)	<30	>90

a. Exonuclease activity was measured on single stranded (³H)T7 DNA. 100% exonuclease activity corresponds to 5,000 units/mg.

45 b. Polymerase activity was measured using single-stranded calf thymus DNA. 100% polymerase activity corresponds to 8,000 units/mg.

Of the seven histidines tested, only one (His 123: mutant 1) has the enzymatic activities characteristic of modified T7 DNA polymerase. T7 gene 5 protein was purified from this mutant using DEAE-cellulose, phosphocellulose, DEAE-Sephadex and hydroxylapatite chromatography. While the polymerase activity was nearly normal (>90% the level of the native enzyme), the exonuclease activity was reduced 4 to 10-fold.

A variant of this mutant was constructed in which both His 123 and Ser 122 were deleted. The gene 5 protein purified from this mutant has a 200-500 fold lower exonuclease activity, again with retention of 55 >90% of the polymerase activity.

These data strongly suggest that His 123 lies in the active site of the exonuclease domain of T7 gene 5 protein. Furthermore, it is likely that the His 123 is in fact the residue being modified by the oxidation involving iron, oxygen and a reducing agent, since such oxidation has been shown to modify histidine

residues in other proteins (Levine, J. Biol. Chem. 258: 11823, 1983; and Hodgson et al. Biochemistry 14: 5294, 1975). The level of residual exonuclease in mutant 11 is comparable to the levels obtainable by chemical modification.

Although mutations at His residues are described, mutations at nearby sites or even at distant sites may also produce mutant enzymes suitable in this invention, e.g., lys and arg (mutants 4 and 15). Similarly, although mutations in some His residues have little effect on exonuclease activity that does not necessarily indicate that mutations near these residues will not affect exonuclease activity.

Mutations which are especially effective include those having deletions of 2 or more amino acids, preferably 6-8, for example, near the His-123 region. Other mutations should reduce exonuclease activity further, or completely.

As an example of the use of these mutant strains the following is illustrative. A pGP5-6 (mutation 11)-containing strain has been deposited with the ATCC (see below). The strain is grown as described above and induced as described in Taber et al. J. Biol. Chem. 262:16212 (1987). K38/pTrx-2 cells may be added to increase the yield of genetically modified T7 DNA polymerase.

The above noted deposited strain also contains plasmid pGP1-2 which expresses T7 RNA polymerase. This plasmid is described in Tabor et al., Proc. Nat. Acad. Sci. USA 82:1074, 1985 and was deposited with the ATCC on March 22, 1985 and assigned the number 40,175.

Referring to Fig. 10, pGP5-6 includes the following segments:

1. EcoRI-SacI-SmaI-BamHI polylinker sequence from M13 mp10 (21bp).

2. T7 bp 14309 to 16747, that contains the T7 gene 5, with the following modifications:

T7 bp 14703 is changed from an A to a G, creating a SmaI site.

T7 bp 14304 to 14321 inclusive are deleted (18 bp).

3. Sall-PstI-HindIII polylinker sequence from M13 mp 10 (15 bp)

4. pBR322 bp 29 (HindIII site) to pBR322 bp 375 (BamHI site).

5. T7 bp 22855 to T7 bp 22927, that contains the T7 RNA Polymerase promoter $\phi 10$, with BamHI linkers inserted at each end (82 bp).

6. pBR322 bp 375 (BamHI site) to pBR322 bp 4361 (EcoRI site).

DNA Sequencing Using Modified T7-type DNA Polymerase

DNA synthesis reactions using modified T7-type DNA polymerase result in chain-terminated fragments of uniform radioactive intensity, throughout the range of several bases to thousands of bases in length. There is virtually no background due to terminations at sites independent of chain terminating agent incorporation (i.e. at pause sites or secondary structure impediments).

Sequencing reactions using modified T7-type DNA polymerase consist of a pulse and chase. By pulse is meant that a short labelled DNA fragment is synthesized; by chase is meant that the short fragment is lengthened until a chain terminating agent is incorporated. The rationale for each step differs from conventional DNA sequencing reactions. In the pulse, the reaction is incubated at 0°C-37°C for 0.5-4 min in the presence of high levels of three nucleotide triphosphates (e.g., dGTP, dCTP and dTTP) and limiting levels of one other labelled, carrier-free, nucleotide triphosphate, e.g., [³⁵S] dATP. Under these conditions the modified polymerase is unable to exhibit its processive character, and a population of radioactive fragments will be synthesized ranging in size from a few bases to several hundred bases. The purpose of the pulse is to radioactively label each primer, incorporating maximal radioactivity while using minimal levels of radioactive nucleotides. In this example, two conditions in the pulse reaction (low temperature, e.g., from 0-20°C, and limiting levels of dATP, e.g., from 0.1μM to 1μM) prevent the modified T7-type DNA polymerase from exhibiting its processive character. Other essential environmental components of the mixture will have similar effects, e.g. limiting more than one nucleotide triphosphate or increasing the ionic strength of the reaction. If the primer is already labelled (e.g., by kinasing) no pulse step is required.

In the chase, the reaction is incubated at 45°C for 1-30 min in the presence of high levels (50-500μM) of all four deoxynucleoside triphosphates and limiting levels (1-50μM) of any one of the four chain terminating agents, e.g., dideoxynucleoside triphosphates, such that DNA synthesis is terminated after an average of 50-600 bases. The purpose of the chase is to extend each radioactively labeled primer under conditions of processive DNA synthesis, terminating each extension exclusively at correct sites in four separate reactions using each of the four dideoxynucleoside triphosphates. Two conditions of the chase (high temperature, e.g., from 30-50°C) and high levels (above 50μM) of all four deoxynucleoside triphosphates) allow the modified T7-type DNA polymerase to exhibit its processive character for tens of thousands of bases; thus the same polymerase molecule will synthesize from the primer-template until a dideoxynucleotide is incorporated. At a chase temperature of 45°C synthesis occurs at >700

nucleotides/sec. Thus, for sequencing reactions the chase is complete in less than a second. ssb increases processivity, for example, when using dTTP, or when using low temperatures or high ionic strength, or low levels of triphosphates throughout the sequencing reaction.

Either [$\alpha^{35}\text{S}$]dATP, [$\alpha^{32}\text{P}$]dATP or fluorescently labelled nucleotides can be used in the DNA sequencing reactions with modified T7-type DNA polymerase. If the fluorescent analog is at the 5' end of the primer, then no pulse step is required.

Two components determine the average extensions of the synthesis reactions. First is the length of time of the pulse reaction. Since the pulse is done in the absence of chain terminating agents, the longer the pulse reaction time, the longer the primer extensions. At 0°C the polymerase extensions average 10 nucleotides/sec. Second is the ratio of deoxyribonucleoside triphosphates to chain terminating agents in the chase reaction. A modified T7-type DNA polymerase does not discriminate against the incorporation of these analogs, thus the average length of extension in the chase is four times the ratio of the deoxynucleoside triphosphate concentration to the chain terminating agent concentration in the chase reaction. Thus, in order to shorten the average size of the extensions, the pulse time is shortened, e.g., to 30 sec. and/or the ratio of chain terminating agent to deoxynucleoside triphosphate concentration is raised in the chase reaction. This can be done either by raising the concentration of the chain terminating agent or lowering the concentration of deoxynucleoside triphosphate. To increase the average length of the extensions, the pulse time is increased, e.g., to 3-4 min; and/or the concentration of chain terminating agent is lowered (e.g., from 20 μM to 2 μM) in the chase reaction.

Example 2: DNA sequencing using modified T7 DNA polymerase

The following is an example of a sequencing protocol using dideoxy nucleotides as terminating agents.

9 μl of single-stranded M13 DNA (mGP1-2, prepared by standard procedures) at 0.7 mM concentration is mixed with 1 μl of complementary sequencing primer (standard universal 17-mer, 0.5 pmole primer / μl) and 2.5 μl 5X annealing buffer (200 mM Tris-HCl, pH 7.5, 50 mM MgCl_2) heated to 65°C for 3 min, and slow cooled to room temperature over 30 min. In the pulse reaction, 12.5 μl of the above annealed mix was mixed with 1 μl dithiothreitol 0.1 M, 2 μl of 3 dNTPs (dGTP, dCTP, dTTP) 3 mM each (P-L Biochemicals in TE), 2.5 μl [$\alpha^{35}\text{S}$]dATP, (1500 Ci/mmol, New England Nuclear) and 1 μl of modified T7 DNA polymerase described in Example 1 (0.4 mg/ml, 2500 units/ml, i.e. 0.4 μg , 2.5 units) and incubated at 0°C, for 2 min, after vortexing and centrifuging in a microfuge for 1 sec. The time of incubation can vary from 30 sec to 20 min and temperature can vary from 0°C to 37°C. Longer times are used for determining sequences distant from the primer.

4.5 μl aliquots of the above pulse reaction are added to each of four tubes containing the chase mixes, preheated to 45°C. The four tubes, labeled G, A, T, C, each contain trace amounts of either dideoxy (dd) G, A, T, or C (P-L Biochemicals). The specific chase solutions are given below. Each tube contains 1.5 μl dATP 1mM, 0.5 μl 5X annealing buffer (200 mM Tris-HCl, pH 7.5, 50mM MgCl_2), and 1.0 μl ddNTP 100 μM (where ddNTP corresponds to ddG,A,T or C in the respective tubes). Each chase reaction is incubated at 45°C (or 30°C-50°C) for 10 min, and then 6 μl of stop solution (90% formamide, 10mM EDTA, 0.1% xylene cyanol) is added to each tube, and the tube placed on ice. The chase times can vary from 1-30 min.

The sequencing reactions are run on standard, 6% polyacrylamide sequencing gel in 7M urea, at 30 Watts for 6 hours. Prior to running on a gel the reactions are heated to 75°C for 2 min. The gel is fixed in 10% acetic acid, 10% methanol, dried on a gel dryer, and exposed to Kodak OM1 high-contrast autoradiography film overnight.

Example 3: DNA sequencing using limiting concentrations of dNTPS

In this example DNA sequence analysis of mGP1-2 DNA is performed using limiting levels of all four deoxyribonucleoside triphosphates in the pulse reaction. This method has a number of advantages over the protocol in example 2. First, the pulse reaction runs to completion, whereas in the previous protocol it was necessary to interrupt a time course. As a consequence the reactions are easier to run. Second, with this method it is easier to control the extent of the elongations in the pulse, and so the efficiency of labeling of sequences near the primer (the first 50 bases) is increased approximately 10-fold.

7 μl of 0.75 mM single-stranded M13 DNA (mGP1-2) was mixed with 1 μl of complementary sequencing primer (17-mer, 0.5 pmole primer/ μl) and 2 μl 5X annealing buffer (200 mM Tris-HCl pH 7.5, 50 mM MgCl_2 , 250 mM NaCl) heated at 65°C for 2 min, and slowly cooled to room temperature over 30 min. In the pulse reaction 10 μl of the above annealed mix was mixed with 1 μl dithiothreitol 0.1 M, 2 μl of 3 dNTPs (dGTP, dCTP, dTTP) 1.5 μM each, 0.5 μl [$\alpha^{35}\text{S}$]dATP, ($\approx 10\mu\text{M}$) (about 10 μM , 1500 Ci/mmol, New England

Nuclear) and 2 μ l modified T7 DNA polymerase (0.1 mg/ml, 1000 units/ml, i.e., 0.2 μ g, 2 units) and incubated at 37°C for 5 min. (The temperature and time of incubation can be varied from 20°C-45°C and 1-60 min., respectively.)

3.5 μ l aliquots of the above pulse reaction were added to each of four tubes containing the chase mixes, which were preheated to 37°C. The four tubes, labeled G, A, T, C, each contain trace amounts of either dideoxy G, A, T, C. The specific chase solutions are given below. Each tube contains 0.5 μ l 5X annealing buffer (200 mM Tris-HCl pH 7.5, 50 mM MgCl₂, 250 mM NaCl), 1 μ l 4dNTPs (dGTP, dATP, dTTP, dCTP) 200 μ M each, and 1.0 μ l ddNTP 20 μ M. Each chase reaction is incubated at 37°C for 5 min (or 20°C-45°C and 1-60 min respectively), and then 4 μ l of a stop solution (95% formamide, 20 mM EDTA, 0.05% xylene-cyanol) added to each tube, and the tube placed on ice prior to running on a standard polyacrylamide sequencing gel as described above.

Example 4: Replacement of dGTP with dITP for DNA sequencing

In order to sequence through regions of compression in DNA, i.e., regions having compact secondary structure, it is common to use dITP (Mills et al., 76 Proc. Natl. Acad. Sci. 2232, 1979) or deazaguanosine triphosphate (deaza GTP, Mizusawa et al., 14 Nuc. Acid Res. 1319, 1986). We have found that both analogs function well with T7-type polymerases, especially with dITP in the presence of ssb. Preferably these reactions are performed with the above described genetically modified T7 polymerase, or the chase reaction is for 1-2 min., and/or at 20°C to reduce exonuclease degradation.

Modified T7 DNA polymerase efficiently utilizes dITP or deaza-GTP in place of dGTP. dITP is substituted for dGTP in both the pulse and chase mixes at a concentration two to five times that at which dGTP is used. In the ddG chase mix ddGTP is still used (not ddITP).

The chase reactions using dITP are sensitive to the residual low levels (about 0.01 units) of exonuclease activity. To avoid this problem, the chase reaction times should not exceed 5 min when dITP is used. It is recommended that the four dITP reactions be run in conjunction with, rather than to the exclusion of, the four reactions using dGTP. If both dGTP and dITP are routinely used, the number of required mixes can be minimized by: (1) Leaving dGTP and dITP out of the chase mixes, which means that the four chase mixes can be used for both dGTP and dITP chase reactions. (2) Adding a high concentration of dGTP or dITP (2 μ l at 0.5 mM and 1-2.5 mM respectively) to the appropriate pulse mix. The two pulse mixes then each contain a low concentration of dCTP, dTTP and [α -³⁵S]dATP, and a high concentration of either dGTP or dITP. This modification does not usually adversely effect the quality of the sequencing reactions, and reduces the required number of pulse and chase mixes to run reactions using both dGTP and dITP to six.

The sequencing reaction is as for example 3, except that two of the pulse mixes contain a) 3 dNTP mix for dGTP: 1.5 μ M dCTP, dTTP, and 1 mM dGTP and b) 3 dNTP mix for dITP: 1.5 μ M dCTP, dTTP, and 2 mM dITP. In the chase reaction dGTP is removed from the chase mixes (i.e. the chase mixes contain 30 μ M dATP, dTTP and dCTP, and one of the four dideoxynucleotides at 8 μ M), and the chase time using dITP does not exceed 5 min.

Deposits

Strains K38/pGP5-5/pTrx-2, K38/pTrx-2 and M13 mGP1-2 have been deposited with the ATCC and assigned numbers 67,287, 67,286, and 40,303 respectively. These deposits were made on January 13, 1987. Strain K38/pGP1-2/pGP5-6 was deposited with the ATCC. On December 4, 1987, and assigned the number 67571.

Applicants' and their assignees acknowledge their responsibility to replace these cultures should they die before the end of the term of a patent issued hereon, 5 years after the last request for a culture, or 30 years, whichever is the longer, and its responsibility to notify the depository of the issuance of such a patent, at which time the deposits will be made irrevocably available to the public. Until that time the deposits will be made irrevocably available to the Commissioner of Patents under the terms of 37 CFR Section 1-14 and 35 USC Section 112.

Claims:

1. A method of producing a purified modified DNA polymerase which method comprises expressing a modified gene which gene encodes a modified processive DNA polymerase which has sufficient DNA polymerase activity for use in DNA sequencing when said polymerase is combined with any cofactor necessary for said DNA polymerase activity and which results from the modification of a naturally

occurring gene modified in that one or more amino acids in the 3' - 5' exonuclease domain of said naturally occurring DNA polymerase are replaced by an amino acid other than that naturally occurring at the site of substitution or are deleted so as to reduce the activity of naturally occurring 3' - 5' exonuclease activity of the naturally occurring DNA polymerase.

- 5 2. A method according to claim 1 characterised in that the polymerase activity of the modified DNA polymerase is at least 90% of that of the naturally occurring DNA polymerase.
- 10 3. A method according to claim 1 or claim 2 further characterised in that said polymerase is a modified bacteriophage T7-type DNA polymerase which has a 3' - 5' exonuclease activity at least 50% lower than the naturally-occurring exonuclease activity of naturally occurring T7-type DNA polymerase.
- 15 4. A method according to any of claims 1 to 3 characterised in that said gene encodes a processive modified DNA polymerase modified to reduce the activity of the naturally occurring 3' - 5' exonuclease activity to less than 500 units per milligram of polymerase.
- 20 5. A method according to any one of claims 1 to 4 further characterised in that said gene has been modified to eliminate the naturally occurring exonuclease activity of the naturally occurring DNA polymerase.
- 25 6. A method according to claim 4 wherein a naturally occurring His residue of the naturally occurring DNA polymerase is replaced or deleted.
7. A method according to any of claims 4 to 6 characterised in that said processive modified T7-type DNA polymerase is T7 DNA polymerase.
8. A method according to claim 7 characterised in that His 123 of the naturally occurring T7 DNA polymerase is replaced or deleted.
- 30 9. A method according to claim 7 characterised in that Ser 122 and His 123 are replaced or deleted.
10. A method according to claim 7 characterised in that amino acid residues Lys 118 to His 123 are deleted.
- 35 11. A method according to claim 7 or claim 9 characterised in that Lys 118 and Arg 119 of the naturally occurring T7 DNA polymerase are replaced or deleted.
- 40 12. A method accord claim 7 characterised in that Arg 131, Lys 136, Lys 140, Lys 144 and Arg 145 of naturally occurring T7 DNA polymerase are replaced or deleted.
13. A method according to any of claims 1 to 12 wherein the modified processive DNA polymerase is able to remain bound to DNA for at least 500 bases under conditions normally used for DNA sequencing reactions.
- 45 14. A purified modified gene characterised in that it encodes a processive modified T7-type DNA polymerase which polymerase is able to remain bound to DNA for at least 500 bases under conditions normally used for DNA sequencing reactions and which has sufficient DNA polymerase activity for use in DNA sequencing when said polymerase is combined with any host factor necessary for said DNA polymerase activity and which results from the modification of a naturally occurring gene modified to reduce the activity of naturally occurring 3' - 5' exonuclease activity of the naturally occurring DNA polymerase wherein one or more amino acids of the exonuclease domain within the amino terminal half of the T7 DNA polymerase of said naturally occurring DNA polymerase, or the corresponding domain of other T7-type DNA polymerases, are replaced by an amino acid other than that naturally occurring at the site of substitution or are deleted.
- 55 15. A purified modified gene according to claim 14 characterised in that one or more of the amino acids of the exonuclease domain from the amino terminal to amino acid residue 224 of T7 DNA polymerase of said naturally occurring DNA polymerase, or the corresponding domain of other T7-type DNA poly-

merases, are replaced by an amino acid other than that naturally occurring at the site of substitution or are deleted.

16. A purified modified gene according to claims 14 or 15 characterised in that the polymerase activity of the processive T7-type modified polymerase is at least 90% of that of the naturally occurring T7-type DNA polymerase.
17. A purified modified gene according to any of claims 14 to 15 further characterised in that said polymerase has an exonuclease activity at least 50% lower than the naturally-occurring exonuclease activity of naturally occurring T7-type DNA polymerase.
18. A purified modified gene according to any of claims 14 to 17 characterised in that said gene encodes a processive modified DNA polymerase modified to reduce the activity of the naturally occurring exonuclease activity to less than 500 units per milligram of polymerase.
19. A purified modified gene as claimed in any of claims 14 to 18 further characterised in that said gene has been modified to eliminate the naturally occurring exonuclease activity of the naturally occurring DNA polymerase.
20. A purified modified gene according to any of claims 14 to 19 further characterised in that a His residue of the 3'- 5' exonuclease domain of the naturally occurring DNA polymerase is replaced or deleted.
21. A purified modified gene according to any of claims 14 to 20 characterised in that said processive modified T7-type DNA polymerase is T7 DNA polymerase.
22. A purified modified gene as claimed in claim 21 characterised in that His 123 of the naturally occurring T7 DNA polymerase is replaced or deleted.
23. A purified modified gene according to claim 21 characterised in that Ser 122 and His 123 are replaced or deleted.
24. A purified modified gene according to claim 21 characterized in that amino acid residues Lys 118 to His 123 are deleted.
25. A purified modified gene according to claim 21 characterized in that Lys 118 and Arg 119 of the naturally occurring T7 DNA polymerase are replaced or deleted.
26. A purified modified gene according to claim 21 characterized in that Arg 131, Lys 136, Lys 140, Lys 144 and Arg 145 of the naturally occurring T7 DNA polymerase are replaced or deleted.
27. The use of a modified processive DNA polymerase produced according to the method of claim 1 for DNA sequencing.

Patentansprüche

1. Verfahren zum Herstellen einer gereinigten, modifizierten DNA-Polymerase, bei dem ein modifiziertes Gen expremiert wird, das für eine modifizierte prozessive DNA-Polymerase codiert, die eine ausreichende DNA-Polymeraseaktivität zur Verwendung beim DNA-Sequenzieren aufweist, wenn die Polymerase mit einem Kofaktor kombiniert wird, der für die DNA-Polymeraseaktivität notwendig ist, und das durch die Modifikation eines natürlich auftretenden Gens erhalten wird, das dahingehend modifiziert ist, daß eine oder mehrere Aminosäuren in der 3'- 5'-Exonukleasedomäne der natürlich auftretenden DNA-Polymerase entfernt oder durch eine Aminosäure ersetzt ist bzw.sind, die sich von der an der Substitutionsstelle natürlich auftretenden Aminosäure unterscheidet, so daß die natürliche Aktivität der 3'-5'-Exonukleaseaktivität der natürlich auftretenden Polymerase vermindert ist.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß die Polymeraseaktivität der modifizierten DNA-Polymerase wenigstens 90% der der natürlich auftretenden DNA-Polymerase ist.

3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß die Polymerase eine modifizierte DNA-Polymerase von Bakteriophagen vom T7-Typ ist, die eine 3'-5'-Exonukleaseaktivität aufweist, die wenigstens 50% niedriger ist als die natürlich auftretende Exonukleaseaktivität der natürlich auftretenden T7-Typ-DNA-Polymerase.
- 5 4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß das Gen für eine prozessive modifizierte DNA-Polymerase codiert, die so modifiziert ist, daß die Aktivität der natürlich auftretenden 3'-5'-Exonukleaseaktivität auf weniger als 500 Einheiten pro Milligramm Polymerase vermindert ist.
- 10 5. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß das Gen modifiziert ist, um die natürlich auftretende Exonukleaseaktivität der natürlich auftretenden DNA-Polymerase zu eliminieren.
- 15 6. Verfahren nach Anspruch 4, bei dem der natürlich auftretende His-Baustein der natürlich auftretenden DNA-Polymerase ersetzt oder entfernt ist.
7. Verfahren nach einem der Ansprüche 4 bis 6, dadurch gekennzeichnet, daß die modifizierte T7-Typ-DNA-Polymerase eine T7-Polymerase ist.
- 20 8. Verfahren nach Anspruch 7, dadurch gekennzeichnet, daß His 123 der natürlich auftretenden T7-DNA-Polymerase ersetzt oder entfernt ist.
9. Verfahren nach Anspruch 7, dadurch gekennzeichnet, daß Ser 122 und His 123 ersetzt oder entfernt sind.
- 25 10. Verfahren nach Anspruch 7, dadurch gekennzeichnet, daß die Aminosäurebausteine Lys 118 bis His 123 entfernt sind.
- 30 11. Verfahren nach Anspruch 7 oder 9, dadurch gekennzeichnet, daß Lys 118 und Arg 119 der natürlich auftretenden T7-DNA-Polymerase ersetzt oder entfernt sind.
12. Verfahren nach Anspruch 7, dadurch gekennzeichnet, daß Arg 131, Lys 136, Lys 140, Lys 144 und Arg 145 der natürlich auftretenden T7-DNA-Polymerase ersetzt oder entfernt sind.
- 35 13. Verfahren nach einem der Ansprüche 1 bis 12, bei dem die modifizierte prozessive DNA-Polymerase unter Bedingungen, wie sie normalerweise für die DNA-Sequenzierungsreaktionen verwendet werden, in der Lage ist, wenigstens 500 Basen lang an der DNA gebunden zu bleiben.
- 40 14. Gereinigtes modifiziertes Gen, dadurch gekennzeichnet, daß es für eine prozessive modifizierte T7-Typ-DNA-Polymerase codiert, die in der Lage ist, unter Bedingungen, die normalerweise für die DNA-Sequenzierungsreaktionen verwendet werden, wenigstens 500 Basen lang an DNA gebunden zu bleiben, und die zur Verwendung bei der DNA-Sequenzierung eine ausreichende DNA-Polymeraseaktivität aufweist, wenn die Polymerase mit einem Wirtsfaktor zusammengebracht wird, der für die Polymeraseaktivität notwendig ist, und das sich aus der Modifikation eines natürlich auftretenden Gens ergibt, das modifiziert ist, um die Aktivität der natürlich auftretenden 3'-5'-Exonukleaseaktivität der natürlich auftretenden Polymerase zu vermindern, bei der eine oder mehrere Aminosäuren der Exonukleasedomäne innerhalb der aminoterminalen Hälfte der T7-DNA-Polymerase der natürlich auftretenden DNA-Polymerase oder die entsprechende Domäne anderer T7-Typ-DNA-Polymerasen fehlt bzw. fehlen oder durch eine Aminosäure ersetzt ist bzw. sind, die sich von der an der Ersatzstelle natürlicherweise vorhandenen Aminosäure unterscheidet.
- 45 50 15. Gereinigtes modifiziertes Gen nach Anspruch 14, dadurch gekennzeichnet, daß eine oder mehrere der Aminosäuren der Exonukleasedomäne von dem Aminoende zu dem Aminosäurebaustein 224 der T7-DNA-Polymerase der natürlich auftretenden DNA-Polymerase oder der entsprechenden Domäne anderer T7-Typ-DNA-Polymerasen entfernt oder durch eine Aminosäure ersetzt ist bzw. sind, die sich von der an der Substitutionstelle natürlich auftretenden Aminosäure unterscheidet.
- 55

16. Gereinigtes modifiziertes Gen nach Anspruch 14 oder 15, dadurch gekennzeichnet, daß die Polymeraseaktivität der prozessiven modifizierten T7-Typ-Polymerase wenigstens 90% der natürlich auftretenden T7-Typ-DNA-Polymerase ist.
- 5 17. Gereinigtes modifiziertes Gen nach einem der Ansprüche 14 und 15, dadurch gekennzeichnet, daß die Polymerase eine Exonukleaseaktivität aufweist, die wenigstens 50% kleiner ist als die natürlich auftretende Exonukleaseaktivität der natürlich auftretenden T7-Typ-DNA-Polymerase.
- 10 18. Gereinigtes modifiziertes Gen nach einem der Ansprüche 14 bis 17, dadurch gekennzeichnet, daß das Gen für eine prozessive modifizierte DNA-Polymerase codiert, die modifiziert ist, um die Aktivität der natürlich auftretenden Exonukleaseaktivität auf weniger als 500 Einheiten pro Milligramm Polymerase zu verringern.
- 15 19. Gereinigtes modifiziertes Gen nach einem der Ansprüche 14 bis 18, dadurch gekennzeichnet, daß das Gen modifiziert wurde, um die natürlich auftretende Exonukleaseaktivität zu eliminieren.
- 20 20. Gereinigtes modifiziertes Gen nach einem der Ansprüche 14 bis 19, dadurch gekennzeichnet, daß ein His-Baustein der 3'-5'-Exonukleasedomäne der natürlich auftretenden DNA-Polymerase ersetzt oder entfernt ist.
21. Gereinigtes modifiziertes Gen nach einem der Ansprüche 14 bis 20, dadurch gekennzeichnet, daß die prozessive modifizierte T7-Typ-DNA-Polymerase eine T7-DNA-Polymerase ist.
- 25 22. Gereinigtes modifiziertes Gen nach Anspruch 21, dadurch gekennzeichnet, daß His 123 der natürlich auftretenden T7-DNA-Polymerase ersetzt oder entfernt ist.
23. Gereinigtes modifiziertes Gen nach Anspruch 21, dadurch gekennzeichnet, daß Ser 122 und His 123 ersetzt oder entfernt sind.
- 30 24. Gereinigtes modifiziertes Gen nach Anspruch 21, dadurch gekennzeichnet, daß die Aminosäurebausteine Lys 118 bis His 123 entfernt sind.
25. Gereinigtes modifiziertes Gen nach Anspruch 21, dadurch gekennzeichnet, daß Lys 118 und Arg 119 der natürlich auftretenden T7-DNA-Polymerase ersetzt oder entfernt sind.
- 35 26. Gereinigtes modifiziertes Gen nach Anspruch 21, dadurch gekennzeichnet, daß Arg 131, Lys 136, Lys 140, Lys 144 und Arg 145 der natürlich auftretenden T7-DNA-Polymerase ersetzt oder entfernt sind.
- 40 27. Verwendung einer modifizierten prozessiven DNA-Polymerase, die nach dem Verfahren nach Anspruch 1 hergestellt ist, zum Sequenzieren.

Revendications

- 45 1. Procédé de production d'une ADN polymérase modifiée et purifiée, selon lequel on exprime un gène modifié, ce gène codant pour une ADN polymérase progressive modifiée ayant une activité d'ADN polymérase suffisante pour être employée dans un séquençage d'ADN lorsque cette polymérase est associée à tout co-facteur nécessaire à cette activité d'ADN polymérase, celui-ci résultant de la modification d'un gène naturel modifié par remplacement d'un ou plusieurs aminoacides dans le domaine d'exonuclease 3'-5' de cette ADN polymérase naturelle, par un aminoacide autre que celui se trouvant naturellement au site de substitution, ou par délétion de ceux-ci de façon à réduire l'activité d'exonuclease naturelle 3'-5' de l'ADN polymérase naturelle.
- 50 2. Procédé selon la revendication 1, caractérisé en ce que l'activité de polymérase de l'ADN polymérase modifiée, est d'au moins 90 % de celle de l'ADN polymérase naturelle.
- 55 3. Procédé selon la revendication 1 ou 2, caractérisé en outre en ce que la polymérase est une ADN polymérase bactériophagique de type T7 modifiée, ayant une activité d'exonuclease 3'-5', au moins 50 % inférieure à l'activité d'exonuclease naturelle de l'ADN polymérase de type T7 naturelle.

4. Procédé selon l'une quelconque des revendications 1 à 3, caractérisé en ce que le gène code pour une ADN polymérase progressive modifiée de façon à réduire l'activité d'exonucléase naturelle 3-5', à moins de 500 unités par mg de polymérase.
5. Procédé selon l'une quelconque des revendications 1 à 4, caractérisé en outre en ce que le gène a été modifié de façon à supprimer l'activité d'exonucléase naturelle de l'ADN polymérase naturelle.
6. Procédé selon la revendication 4, dans lequel on remplace ou on supprime un résidu His naturel de l'ADN polymérase naturelle.
7. Procédé selon l'une quelconque des revendications 4 à 6, caractérisé en ce que l'ADN polymérase progressive modifiée de type T7 est une ADN polymérase T7.
8. Procédé selon la revendication 7, caractérisé en ce que le résidu His 123 de l'ADN polymérase T7 naturelle, est remplacé ou supprimé.
9. Procédé selon la revendication 7, caractérisé en ce que les résidus Ser 122 et His 123 sont remplacés ou supprimés.
10. Procédé selon la revendication 7, caractérisé en ce que les résidus dérivés d'acide aminé Lys 118 à His 123, sont supprimés.
11. Procédé selon la revendication 7 ou 9, caractérisé en ce que les résidus Lys 118 et Arg 119 de l'ADN polymérase T7 naturelle, sont remplacés ou supprimés.
12. Procédé selon la revendication 7, caractérisé en ce que les résidus Arg 131, Lys 136, Lys 140, Lys 144 et Arg 145 de l'ADN polymérase T7 naturelle, sont remplacés ou supprimés.
13. Procédé selon l'une quelconque des revendications 1 à 12, dans lequel l'ADN polymérase progressive modifiée, est capable de rester liée à de l'ADN sur au moins 500 bases dans les conditions normalement employées pour des réactions de séquençage de l'ADN.
14. Gène modifié et purifié, caractérisé en ce qu'il code pour une ADN polymérase progressive modifiée de type T7, cette polymérase étant capable de rester liée à de l'ADN sur au moins 500 bases dans les conditions normalement employées pour des réactions de séquençage d'ADN, et ayant une activité d'ADN polymérase suffisante pour une utilisation dans un séquençage d'ADN lorsque cette polymérase est associée à n'importe quel facteur d'hôte nécessaire à cette activité d'ADN polymérase, résultant de la modification d'un gène naturel, modifié de façon à réduire l'activité d'exonucléase naturelle 3'-5' de l'ADN polymérase naturelle, un ou plusieurs acides aminés du domaine d'exonucléase compris dans la moitié amino-terminale de la polymérase, ou dans le domaine correspondant d'autres ADN polymérases de type T7, étant remplacés par un acide aminé autre que celui naturellement présent au site de substitution, ou ceux-ci étant supprimés.
15. Gène modifié et purifié selon la revendication 14, caractérisé en ce qu'un ou plusieurs des acides aminés du domaine d'exonucléase allant de l'extrémité amino-terminale jusqu'au résidu dérivé d'acide aminé 224 de l'ADN polymérase T7 naturelle, ou du domaine correspondant d'autres ADN polymérases de type T7 sont remplacés par un acide aminé autre que celui naturellement présent dans le site de substitution, ou ils sont supprimés.
16. Gène modifié et purifié selon la revendication 14 ou 15, caractérisé en ce que l'activité de polymérase de la polymérase progressive modifiée de type T7 est d'au moins 90 % celle de l'ADN polymérase naturelle de type T7.
17. Gène modifié et purifié selon la revendication 14 ou 15, caractérisé en ce qu'en outre la polymérase a une activité exonucléase, au moins 50 % inférieure à l'activité d'exonucléase naturelle de l'ADN polymérase naturelle de type T7.

18. Gène modifié et purifié selon l'une quelconque des revendications 14 à 17, caractérisé en ce que le gène code pour une ADN polymérase progressive et modifiée de façon à réduire l'activité d'exonucléase naturelle, à moins de 500 unités par mg de polymérase.
- 5 19. Gène modifié et purifié selon l'une quelconque des revendications 14 à 18, caractérisé en outre en ce que ce gène a été modifié de façon à supprimer l'activité d'exonucléase naturelle de l'ADN polymérase naturelle.
20. Gène modifié et purifié selon l'une quelconque des revendications 14 à 19, caractérisé en outre en ce qu'un résidu His du domaine d'exonucléase 3-5' de l'ADN polymérase naturelle, est remplacé ou supprimé.
- 10 21. Gène modifié ou purifié selon l'une quelconque des revendications 14 à 20, caractérisé en ce que cette ADN polymérase progressive modifiée de type T7 est une ADN polymérase de type T7.
- 15 22. Gène modifié et purifié selon la revendication 21, caractérisé en ce que le résidu His 123 de l'ADN polymérase T7 naturelle est remplacé ou supprimé.
- 20 23. Gène modifié et purifié selon la revendication 21, caractérisé en ce que les résidus Ser 122 et His 123 sont remplacés ou supprimés.
24. Gène modifié et purifié selon la revendication 21, caractérisé en ce que les résidus dérivés d'acide Lys 118 à His 123 sont supprimés.
- 25 25. Gène modifié et purifié selon la revendication 21, caractérisé en ce que les résidus Lys 118 et Arg 119 de l'ADN polymérase T7 naturelle sont remplacés ou supprimés.
- 30 26. Gène modifié et purifié selon la revendication 21, caractérisé en ce que les résidus Arg 131, Lys 136, Lys 140, Lys 144 et Arg 145 de l'ADN polymérase T7 naturelle, sont remplacés ou supprimés.
- 35 27. Utilisation d'une ADN polymérase progressive modifiée, produite selon le procédé de la revendication 1, pour le séquençage d'ADN.

35

40

45

50

55

FIGURE 1

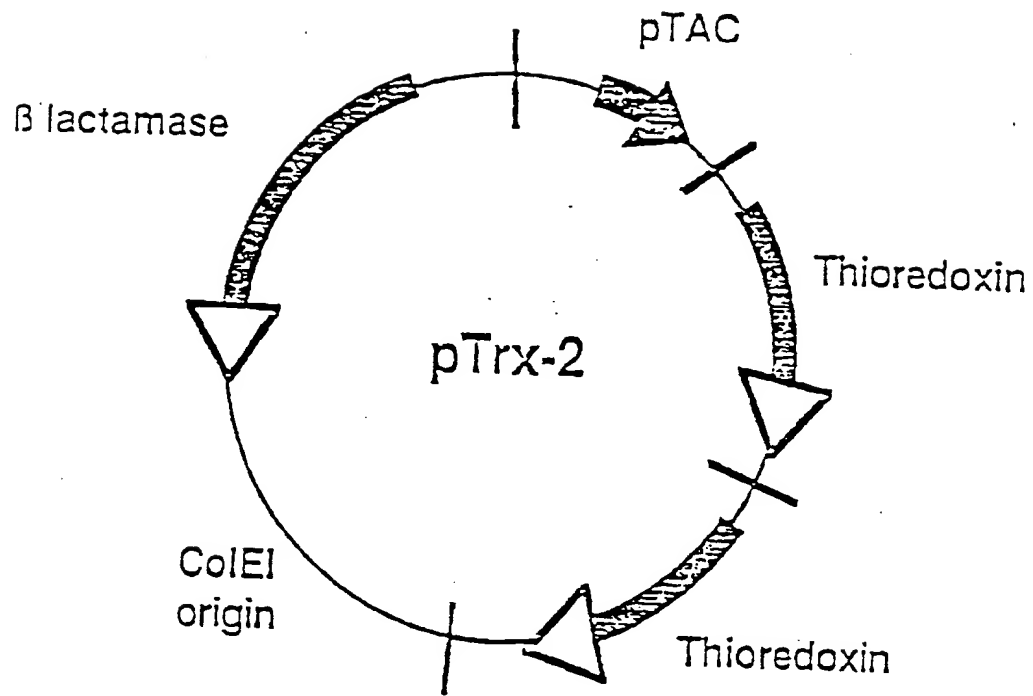


FIGURE 2

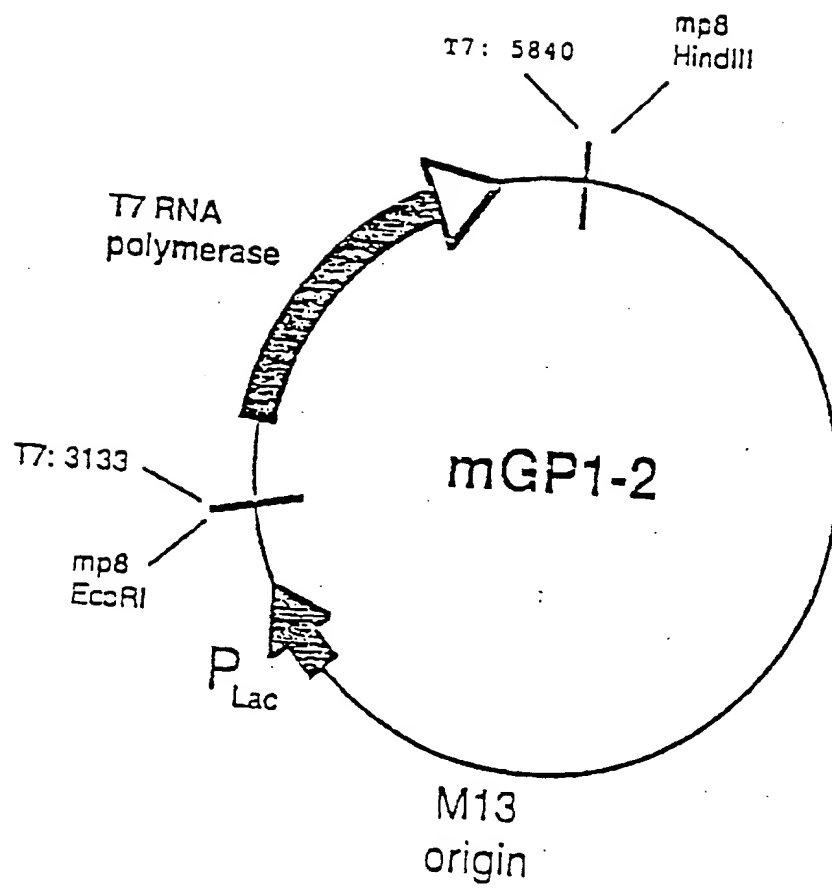


FIGURE 3

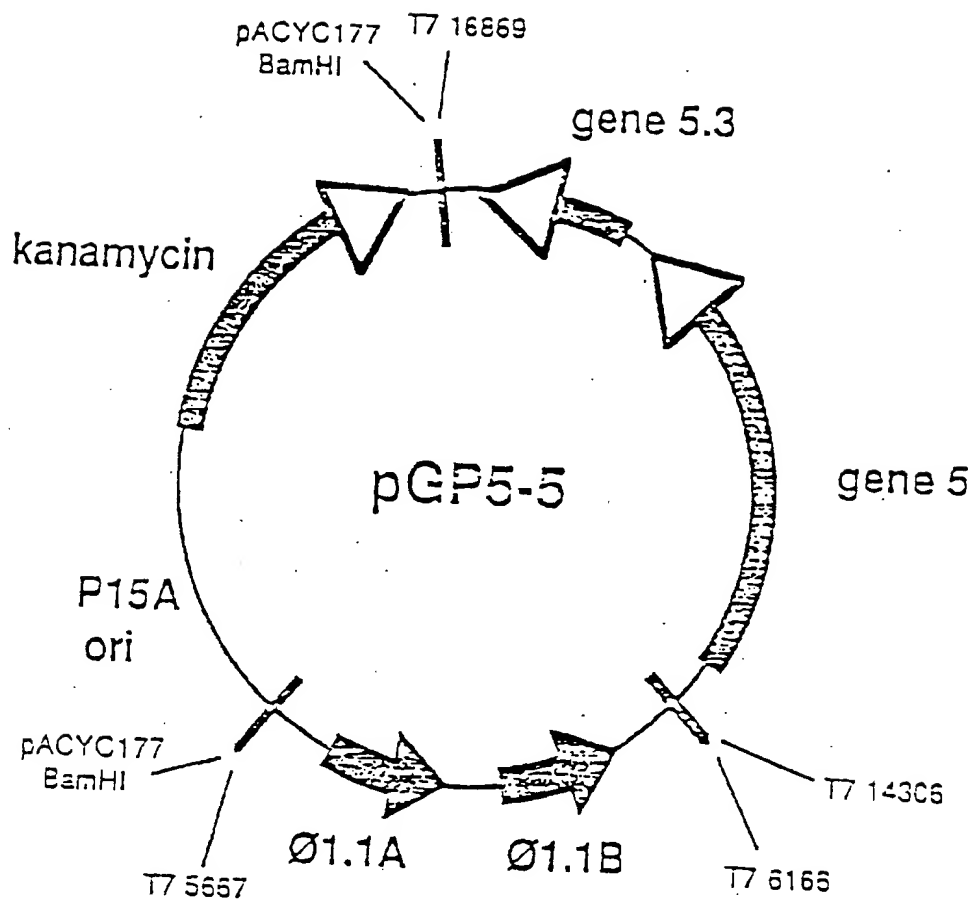


FIGURE 4

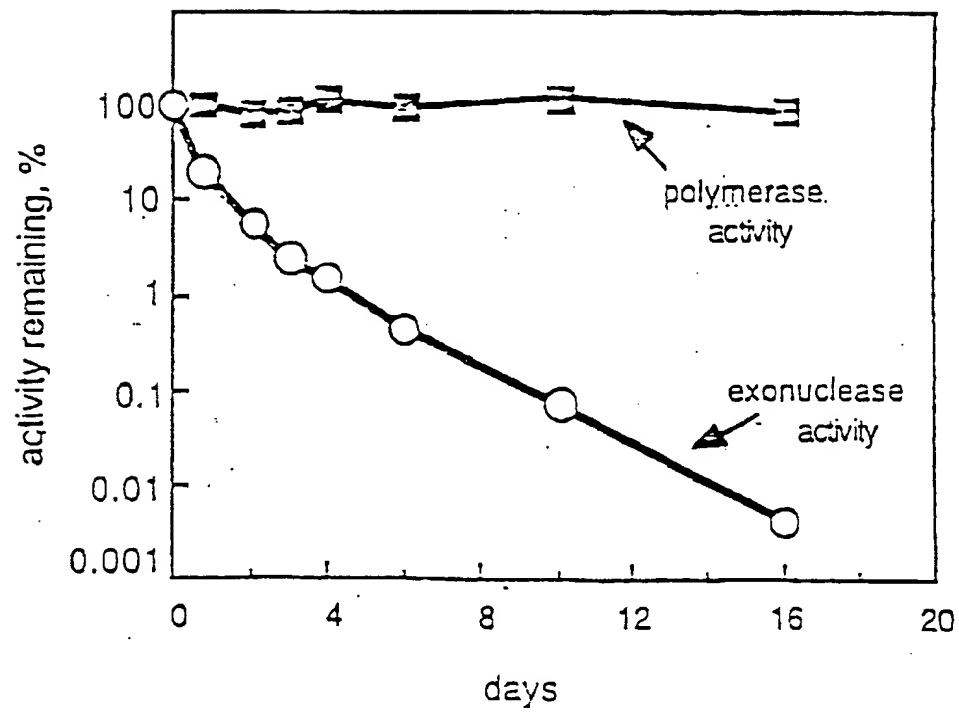


FIGURE 5

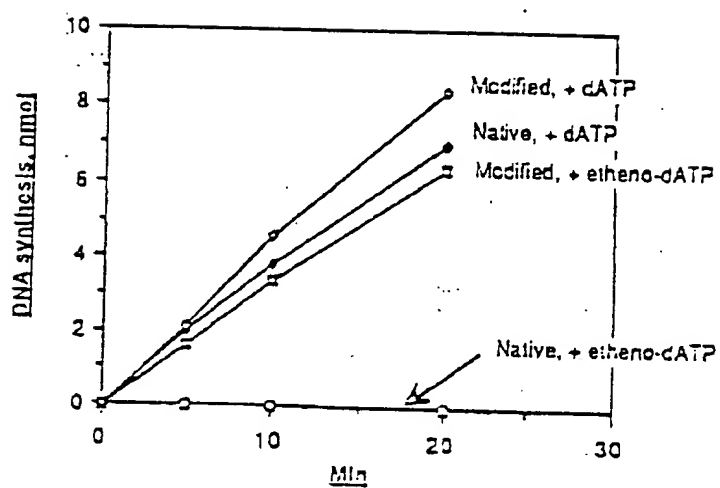


FIGURE 6

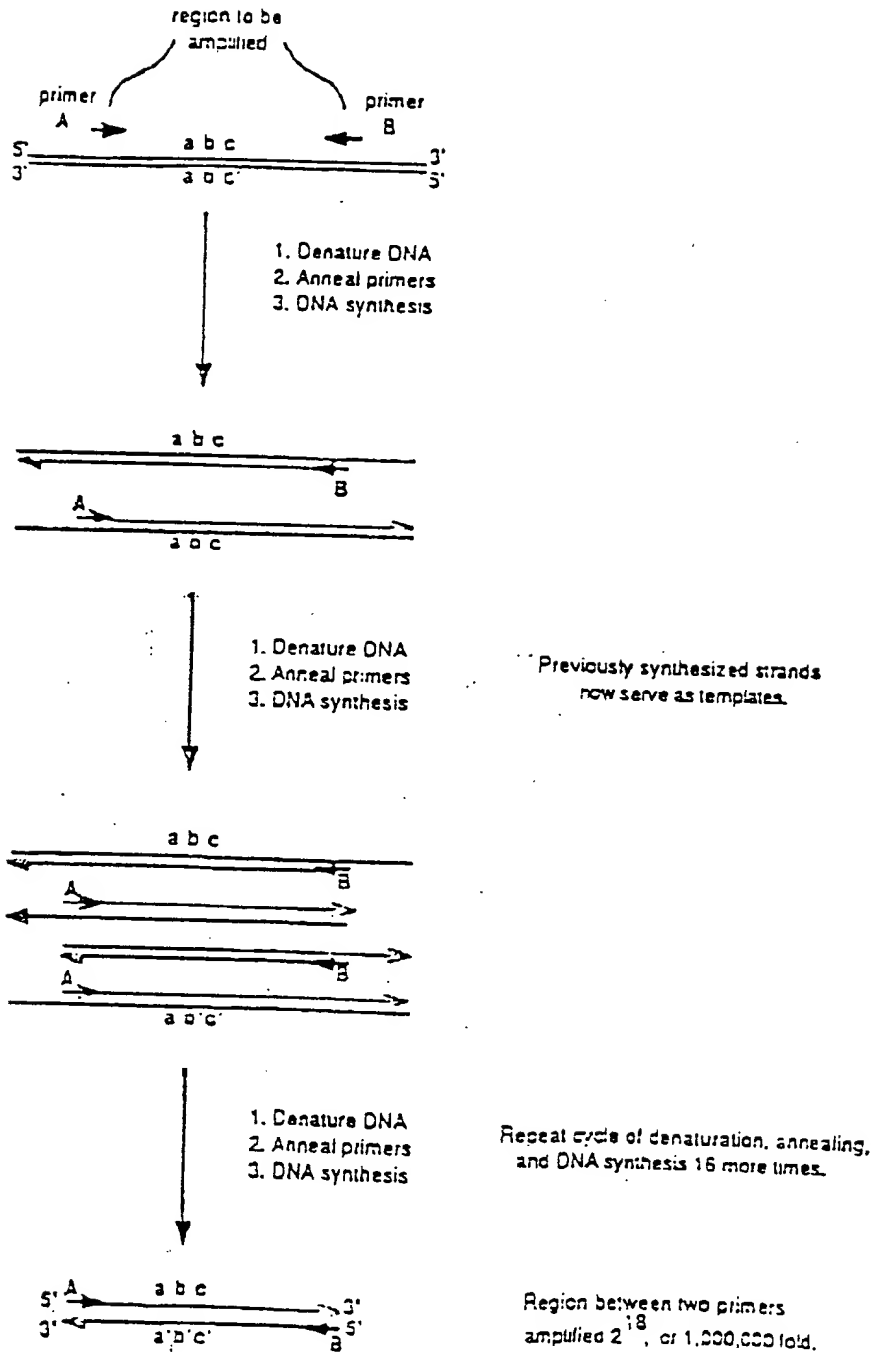


FIGURE 7

10	20	30	40	50
TTCTTCTCAT	GTTTGACAGC	TTATCATCGA	CTGCACGGTG	CACCAATGCT
60	70	80	90	100
TCTGGCGTCA	GGCAGCCATC	GGAAGCTGTG	GTATGGCTGT	GCAGGTCGTA
110	120	130	140	150
AATCACTGCA	TAATTCGTGT	CGCTCAAGGC	GCACTCCCGT	TCTGGATAAT
160	170	180	190	200
GTTTTTTGCG	CCGACATCAT	AACGGTTCTG	GCAAATATTC	TGAAATGAGC
210	220	230	240	250
TGTTGACAAT	TAATCATCGG	CTCGTATAAT	GTGTGGAATT	GTGAGCGGAT
260	270	280	290	300
AACAATTTCA	CACAGGAAAC	AGGGGATCCG	TCAACCTTTA	GTTGGTTAAT
310	320	330	340	350
GTTACACCAA	CAACGAAACC	AACACGCCAG	GCTTATTCCT	GTGGAGTTAT
360	370	380	390	400
ATATGAGCGA	TAAAATTATT	CACCTGACTG	ACGACAGTTT	TGACACGGAT
410	420	430	440	450
GTACTCAAAG	CGGACGGGGC	GATCCTCGTC	GATTTCTGGG	CAGAGTGGTG
460	470	480	490	500
CGGTCCGTGC	AAGATGATCG	CCCCGATTCT	GGATGAAATC	GCTGACGAAT

FIGURE 7 (continued)

510	520	530	540	550
ATCAGGGCAA	ACTGACCGTT	GCAAACTGA	ACATCGATCA	AAACCCCTGGT
560	570	580	590	600
ACTSCGCCGA	AATATGGCAT	CCGTGGTATC	CCGACTCTGC	TGCTGTTCAA
610	620	630	640	650
AAACGGTGAA	GTGGCGGCAA	CCAAAGTGGG	TGCACTGTCT	AAAGGTCAGT
660	670	680	690	700
TGAAGAGATT	CCTCGACGCT	AACCTGGCGT	AAGGGAATTT	CATGTTCCGG
710	720	730	740	750
TGCCCCGTGC	CTAAAACTG	GACGCCCGGC	GTGAGTCATG	CTAACTTAGT
760	770	780	790	800
GTTGACGGAT	CCCCGGGGAT	CCGTCAACCT	TTAGTTGGTT	AATGTTACAC
810	820	830	840	850
CAACAACGAA	ACCAACACGC	CAGGCTTATT	CCTGTGGAGT	TATATATGAG
860	870	880	890	900
CGATAAAATT	ATTCACTGA	CTGACGACAG	TTTTGACACG	GATGTACTCA
910	920	930	940	950
AAGCGGACGG	GGCGATCCTC	GTGATTTTCT	GGGCAGAGTG	GTGCGGTCCG
960	970	980	990	1000
TGCAAGATGA	TCGCCCCGAT	TCTGGATGAA	ATCGCTGACG	AATATCAGGG
1010	1020	1030	1040	1050
CAAACTGACC	GTTGCAAAAC	TGAACATCGA	TCAAAACCCT	GGTACTGCGC
1060	1070	1080	1090	1100
CGAAATATGG	CATCCGTGGT	ATCCCGACTC	TGCTGCTGTT	CAAAAACGGT
1110	1120	1130	1140	1150
GAAGTGGCGG	CAACCAAGT	GGGTGCACTG	TCTAAAGGTC	AGTTGAAGA
1160	1170	1180	1190	1200
GTTCTCGAC	GCTAACCTGG	CGTAAGGGAA	TTTCATGTTT	GGGTGCCCCG
1210	1220	1230	1240	1250
TCGCTAAAA	CTGGACGCCC	GGCGTGAGTC	ATGCTAACTT	AGTGTTGACG
1260	1270	1280	1290	1300
GATCCCCCTG	CCTCSCCGCT	TTCGGTGATG	ACGGTGAAGA	CCTCTGACAC
1310	1320	1330	1340	1350
ATGCAGCTCC	CGGAGACGGT	CACAGCTTGT	CTGTAAGCGG	ATGCCGGGAG
1360	1370	1380	1390	1400
CAGACAAGCC	CGTCAGGGCG	CGTCAGCGGG	TGTTGGCGGG	TGTCGGGGCG
1410	1420	1430	1440	1450
CAGCCATGAC	CCAGTCACGT	AGCGATAGCG	GAGTGTATAC	TGGCTTAACT
1460	1470	1480	1490	1500
ATGCGGCATC	AGAGCAGATT	GTAATGAGAG	TGCACCATAT	GCGGTGTGAA
1510	1520	1530	1540	1550
ATACCGCACA	GATGCGTAAG	GAGAAATATC	CGCATCAGGC	GCTCTTCCGC
1560	1570	1580	1590	1600
TTCCTCSCCTC	ACTGACTCSC	TGCGCTCCTT	CGTTCCGGCTG	CGGCGAGCGG
1610	1620	1630	1640	1650
TATCAGCTCA	CTCAAAGGCG	GTAATACGGT	TATCCACAGA	ATCAGGGGAT
1660	1670	1680	1690	1700
AACGCGAGGAA	AGAACATGTC	AGCAAAAGGC	CAGCAAAAGG	CCAGGAACCG
1710	1720	1730	1740	1750
TAAAAAGGCC	GCCTTGCTGG	CGTTTTTCCA	TAGGCTCCGC	CCCCCTGACC
1760	1770	1780	1790	1800
AGCATCAGAA	AAATCGACGC	TCAAGTCAGA	GGTGCGGAAA	CCCGACAGGA
1810	1820	1830	1840	1850
CTATAAAGAT	ACCAGGCGTT	TCCCCCTGGA	AGCTCCCTCG	TGCGCTCTCC

FIGURE 7 (continued)

1860	1870	1880	1890	1900
TGTTCCGACC	CTSCCGCTTA	CCGGATACCT	GTCCGCCTTT	CTCCCTTCGG
1910	1920	1930	1940	1950
GAAGCGTGGC	GCTTTCTCAA	TGCTCACGCT	GTAGGTATCT	CAGTTCGGTG
1960	1970	1980	1990	2000
TAGGTCGTTC	GCTCCAAGCT	GGGCTGTGTG	CACGAACCCC	CCGTTTCAGCC
2010	2020	2030	2040	2050
CGACCGCTGC	GCCTTATCCG	GTAACATATCG	TCTTGAGTCC	AACCCGGTAA
2060	2070	2080	2090	2100
GACACGACTT	ATCGCCACTG	GCAGCAGCCA	CTGGTAACAG	GATTAGCAGA
2110	2120	2130	2140	2150
GCGAGGIATG	TAGGCGGTGC	TACAGAGTTC	TTGAAGTGGT	GGCCTAACTA
2160	2170	2180	2190	2200
CGGCTACACT	AGAAGGACAG	TATTTGGTAT	CTGCGCTCTG	CTGAAGCCAG
2210	2220	2230	2240	2250
TTACCTTCGG	AAAAAGAGTT	GGTAGCTCTT	GATCCGGCAA	ACAAACCACC
2260	2270	2280	2290	2300
GCTGGTAGCG	GTGGTTTTTT	TGTTTGCAAG	CAGCAGATTA	CGCGCAGAAA
2310	2320	2330	2340	2350
AAAAGGATCT	CAAGAAGATC	CTTTGATCTT	TTCTACGGGG	TCTGACGCTC
2360	2370	2380	2390	2400
AGTGGAACGA	AAACTCACGT	TAAGGGATT	TGGTCATGAG	ATTATCAAAA
2410	2420	2430	2440	2450
AGGATCTTCA	CCTAGATCCT	TTTAAATTAA	AAATGAAGTT	TTAAATCAAT
2460	2470	2480	2490	2500
CTAAAGTATA	TATGAGTAAA	CTTGGTCTGA	CAGTTACCAA	TGCTTAATCA
2510	2520	2530	2540	2550
GTGAGGCACC	TATCTCAGCG	ATCTGTCTAT	TTCGTTCTAT	CATAGTTGCC
2560	2570	2580	2590	2600
TGACTCCCCG	TCGTGTAGAT	AACTACGATA	CGGGAGGGCT	TACCATCTGG
2610	2620	2630	2640	2650
CCCCAGTGCT	GCAATGATAC	CGCGAGACCC	ACGCTCACCG	GCTCCAGATT
2660	2670	2680	2690	2700
TATCAGCAAT	AAACCAGCCA	GCCGGAAGGG	CCGAGCGCAG	AAGTGGTCCT
2710	2720	2730	2740	2750
GCAACTTTAT	CCGCCTCCAT	CCAGTCTATT	AATTGTTGCC	GGGAAGCTAG
2760	2770	2780	2790	2800
AGTAAGTAGT	TCGCCAGTTA	ATAGTTTGCG	CAACGTTGTT	GCCATTGCTG
2810	2820	2830	2840	2850
CAGGCATCGT	GGTGTACCGC	TCGTGTTTTG	GTATGGCTTC	ATTCAGCTCC
2860	2870	2880	2890	2900
GGTTCCCAAC	GATCAAGGCG	AGTTACATGA	TCCCCCATGT	TGTGCAAAAA
2910	2920	2930	2940	2950
AGCGGTTAGC	TCCTTCGGTC	CTCCGATCGT	TGTCAGAAAT	AAGTTGGCCG
2960	2970	2980	2990	3000
CAGTGTTATC	ACTCATGGTT	ATGGCAGCAC	TGCATAATTC	TCTTACTGTC
3010	3020	3030	3040	3050
ATGCCATCCG	TAAGATGCTT	TTCTGTGACT	GGTGAGTACT	CAACCAAGTC
3060	3070	3080	3090	3100
ATTCTGAGAA	TAGTGTATGC	GGCGACCGAG	TTGCTCTTGC	CCGGCGTCAA
3110	3120	3130	3140	3150
CACGGGATAA	TACCGCGCCA	CATAGCAGAA	CTTTAAAAGT	GCTCATCATT
3160	3170	3180	3190	3200
GGAAAACGTT	CTTCGGGGCG	AAACTCTCA	AGGATCTTAC	CGCTSTTSAG

FIGURE 7 (continued)

3210	3220	3230	3240	3250
ATCCAGTTCG	ATGTAACCCA	CTCGTGCACC	CAACTGATCT	TCAGCATCTT
3260	3270	3280	3290	3300
TTACTTTCAC	CAGCGTTTCT	GGGTGAGCAA	AAACAGGAAG	GCAAAATGCC
3310	3320	3330	3340	3350
GCAAAAAAGG	GAATAAGGGC	GACACGGAAA	TGTTGAATAC	TCATACTCTT
3360	3370	3380	3390	3400
CCTTTTTCAA	TATTATTGAA	GCATTTATCA	GGGTTATTGT	CTCATGAGCG
3410	3420	3430	3440	3450
GATACATATT	TGAATGTATT	TAGAAAAATA	AACAAATAGG	GGTTCCGCGC
3460	3470	3480	3490	3500
ACATTTCCCC	GAAAAGTGCC	ACCTGACGTC	TAAGAAACCA	TTATTATCAT
3510	3520	3530	3540	3550
GACATTAACC	TATAAAAATA	GGCGTATCAC	GAGGCCCTTT	CGTCTTCAAG

AA

FIGURE 3

10	20	30	40	50
GTTGACACAT	ATGAGTCTTG	TGATGTA CTG	GCTGATTTCT	ACGACCA GTT
60	70	80	90	100
CGCTGACCAG	TTGCACGAGT	CTCAATTGGA	CAAAATGCCA	GCACTTCCGG
110	120	130	140	150
CTAAAGGTAA	CTTSAACCTC	CGTGACATCT	TAGAGTCGGA	CTTCGCGTTC
160	170	180	190	200
GCGTAACGCC	AAATCAATAC	GACTCACTAT	AGAGGGACAA	ACTCAAGGTC
210	220	230	240	250
ATTCGCAAGA	GTGGCCTTTA	TGATTGACCT	TCTTCCGGTT	AATACGACTC
260	270	280	290	300
ACTATAGGAG	AACCTTAAGG	TTTAACTTTA	AGACCCTTAA	GTGTTAATTA
310	320	330	340	350
GAGATTTAAA	TTAAAGAATT	ACTAAGAGAG	GACTTTAAGT	ATGCGTAACT
360	370	380	390	400
TCGAAAAGAT	GACCAAACGT	TCTAACCGTA	ATGCTCGTGA	CTTCGAGGCA
410	420	430	440	450
ACCAAAGGTC	GCAAGTTGAA	TAAGACTAAG	CGTGACCGCT	CTCACAAGCG
460	470	480	490	500
TAGCTGGGAG	GGTCAGTAAG	ATGGGACGTT	TATATAGTGG	TAATCTGGCA
510	520	530	540	550
CCGGATCCGG	TATGAAGAGA	TTGTTAAGTC	ACGATAATCA	ATAGGAGAAA
560	570	580	590	600
TCAATATGAT	CGTTTCTGAC	ATCGAAGCTA	ACGCCCTCTT	AGAGAGCGTC

FIGURE 8 (continued)

610	620	630	640	650
ACTAAGTTCC	ACTGCGGGGT	TATCTACGAC	TACTCCACCG	CTGAGTACGT
660	670	680	690	700
AAGCTACCGT	CCGAGTGACT	TCGGTGCCTA	TCTGGATGCG	CTGGAAGCCG
710	720	730	740	750
AGGTTGCACG	AGGCGGTCTT	ATTGTGTCTC	ACAACGGTCA	CAAGTATGAC
760	770	780	790	800
GTTCTCTCAT	TGACCAAACT	GGCAAAGTTG	CAATTGAACC	GAGAGTTCCA
810	820	830	840	850
CCTTCCTCGT	GAGAACTGTA	TTGACACCCCT	TGTGTTGTCA	CGTTTGATTC
860	870	880	890	900
ATTCCAACCT	CAAGGACACC	GATATGGGTC	TTCTGCCTTC	CGGCAAGTTG
910	920	930	940	950
CCCGGAAAAC	GCITTTGGGTC	TCACGCTTTG	GAGGCGTGGG	GTATATCGCTT
960	970	980	990	1000
AGGCGAGATG	AAGGGTGAAT	ACAAAGACGA	CTTTAAGCGT	ATGCTTGAAG
1010	1020	1030	1040	1050
AGCAGGGTGA	AGAATACGTT	GACGGAATGG	AGTGGTGGAA	CTTCAACGAA
1060	1070	1080	1090	1100
GAGATGATGG	ACTATAACGT	TCAGGACGTT	GTGGTAACCTA	AAGCTCTCCT
1110	1120	1130	1140	1150
TGAGAAGCTA	CTCTCTGACA	AACATTACTT	CCCTCCTGAG	ATTGACTTTA
1160	1170	1180	1190	1200
CGGACGTAGG	ATACACTACG	TTCTGGTCAG	AATCCCTTGA	GGCCGTTGAC
1210	1220	1230	1240	1250
ATTGAACATC	GTGCTGCATG	GCTGCTCGCT	AAACAAGAGC	GCAACGGGTT
1260	1270	1280	1290	1300
CCCGTTTGAC	ACAAAAGCAA	TCGAAGAGTT	GTACGTAGAG	TTAGCTGCTC
1310	1320	1330	1340	1350
GCCGCTCTGA	GTTGCTCCGT	AAATTGACCG	AAACGTTCCG	CTCGTGGTAT
1360	1370	1380	1390	1400
CAGCCTAAAG	GTGGCACTGA	GATGTTCTGC	CATCCGCGAA	CAGGTAAAGC
1410	1420	1430	1440	1450
ACTACCTAAA	TACCCTCGCA	TTAAGACACC	TAAAGTTGGT	GGTATCTTTA
1460	1470	1480	1490	1500
AGAAGCCTAA	GAACAAGGCA	CAGCGAGAAG	GCCGTGAGCC	TTGCGAAGCTT
1510	1520	1530	1540	1550
GATACCCGCG	AGTACGTTSC	TGGTGCCTCCT	TACACCCGAG	TTGAACATGT
1560	1570	1580	1590	1600
TGTGTTTAAC	CCTTCGTCTC	GTGACCACAT	TCAGAAGAAA	CTCCAAGAGG
1610	1620	1630	1640	1650
CTGGGTGGGT	CCCGACCAAG	TACACCGATA	AGGGTGCTCC	TGTGGTGGAC
1660	1670	1680	1690	1700
GATGAGGTAC	TCGAAGGAGT	ACGTGTAGAT	GACCCCTGAGA	AGCAAGCCCG
1710	1720	1730	1740	1750
TATCGACCTC	ATTAAAGAGT	ACTTGATGAT	TCAGAAGCGA	ATCGGACAGT
1760	1770	1780	1790	1800
CTGCTGAGGG	AGACAAAGCA	TGGCTTCGTT	ATGTTGCTGA	GGATGGTAAG
1810	1820	1830	1840	1850
ATTGATGGTT	CTGTTAAGCC	TAATGGAGCA	GTTACGGGTC	GTGCGACCCA
1860	1870	1880	1890	1900
TGCGTTCCCA	AACCTTCCGC	AAATTCCGSG	TGTACGTTCT	CCTTATGGAG
1910	1920	1930	1940	1950
AGCAGTGTCS	CGCTGCTTTT	GGCGCTGASC	ACCATTTGGA	TGGGATAACT

FIGURE 8 (continued)

1960	1970	1980	1990	2000
GGTAAGCCTT	GGGTTTCAGGC	TGGCAATCGAC	GCATCCGGTC	TTGAGCTACG
2010	2020	2030	2040	2050
CTGCTTGGCT	CACTTCATGG	CTCGCTTTGA	TAACGGCGAG	TACGCTCAGG
2060	2070	2080	2090	2100
AGATTCTTAA	CGGCGACATC	CACACTAAGA	ACCAGATAGC	TGCTGAACTA
2110	2120	2130	2140	2150
CCTACCCGAG	ATAACGCTAA	GACGTTTCATC	TATGGGTTCC	TCTATGGTGC
2160	2170	2180	2190	2200
TGGTGATGAG	AAGATTGGAC	AGATTGTTGG	TGCTGGTAAA	GAGCGCGGTA
2210	2220	2230	2240	2250
AGGAACTCAA	GAAGAAATTC	CTTGAGAACA	CCCCCGCGAT	TGCAGCACTC
2260	2270	2280	2290	2300
CGCGAGTCTA	TCCAACAGAC	ACTTGTCGAG	TCCTCTCAAT	GGGTAGCTGG
2310	2320	2330	2340	2350
TGAGCAACAA	GTCAAGTGGA	AACGCCGCTG	GATTAAAGGT	CTGGATGGTC
2360	2370	2380	2390	2400
GTAAGGTACA	CGTTCGTAGT	CCTCACGCTG	CCTTGAATAC	CCTACTGCAA
2410	2420	2430	2440	2450
TCTGCTGGTG	CTCTCATCTG	CAAACGTGG	ATTATCAAGA	CCGAAGAGAT
2460	2470	2480	2490	2500
GCTCGTAGAG	AAAGGCTTGA	AGCATGGCTG	GGATGGGGAC	TTTGCCTACA
2510	2520	2530	2540	2550
TGGCATGGGT	ACATGATGAA	ATCCAAGTAG	GCTGCCGTAC	CGAAGAGATT
2560	2570	2580	2590	2600
GCTCAGGTGG	TCATTGAGAC	CGCACAAGAA	GCGATGCGCT	GGGTTGGAGA
2610	2620	2630	2640	2650
CCACTGGAAC	TTCCGGTGTC	TTCTGGATAC	CGAAGGTAAG	ATGGGTCCTA
2660	2670	2680	2690	2700
ATTGGGCGAT	TTGCCACTGA	TACAGGAGGC	TACTCATGAA	CGAAAGACAC
2710	2720	2730	2740	2750
TTAACAGGTG	CTGCTTCTGA	AATGCTAGTA	GCCTACAAAT	TTACCAAAGC
2760	2770	2780	2790	2800
TGGGTAGACT	GTCTATTACC	CTATGCTGAC	TCAGAGTAAA	GAGGACTTGG
2810	2820	2830	2840	2850
TTGTATGTAA	GGATGGTAAA	TTTAGTAAGG	TTCAGGTTAA	AACAGCCACA
2860	2870	2880	2890	2900
ACGGTTCAAA	CCAACACAGG	AGATGCCAAG	CAGGTTAGGC	TAGGTGGATG
2910	2920	2930	2940	2950
CGGTAGGTCC	GAATATAAGG	ATGGAGACTT	TGACATCTTT	GCGGTTGTGG
2960	2970	2980	2990	3000
TTGACGAAGA	TGTGCTTATT	TTACATGGG	ACGAAGTAAA	AGGTAAGACA
3010	3020	3030	3040	3050
TCCATGTGTG	TGGGCAAGAG	AAACAAAGGC	ATAAACTAT	AGGAGAAATT
3060	3070	3080		
ATTATGGCTA	TGACAAAGAA	ATTTCCGGAT	C	

FIGURE 9

10	20	30	40	50
AATGCTACTA	CTATTAGTAG	AATTGATGCC	ACCTTTTCAG	CTCGCGCCCC
60	70	80	90	100
AAATGAAAAT	ATAGCTAAAC	AGGTTATTGA	CCATTGCGA	AATGTATCTA
110	120	130	140	150
ATGGTCAAAC	TAAATCTACT	CGTTCGCAGA	ATTGGGAATC	AACTGTTACA
160	170	180	190	200
TGGAATGAAA	CTTCCAGACA	CCGTACTTTA	GTTGCATATT	TAAAACATGT
210	220	230	240	250
TGAGCTACAG	CACCAGATTC	AGCAATTAAG	CTCTAAGCCA	TCCGCAAAAA
260	270	280	290	300
TGACCTCTTA	TCAAAAGGAG	CAATTAAAGG	TACTCTCTAA	TCCTGACCTG
310	320	330	340	350
TTGGAGTTTG	CTTCCGGTCT	GGTTCGCTTT	GAAGCTCGAA	TTAAAACGCG
360	370	380	390	400
ATATTTGAAG	TCITTCGGGC	TTCTCTTTAA	TCTTTTGTAT	GCAATCCGCT
410	420	430	440	450
TTGCTTCTGA	CTATAATAGT	CAGGGTAAAG	ACCTGATTTT	TGATTTATGG
460	470	480	490	500
TCATTCTCGT	TTTCTGAAT	GTTTAAAGCA	TTTGAGGGGG	ATTCAATGAA
510	520	530	540	550
TATTTATGAC	GATTCGCGAG	TATTGGACGC	TATCCAGTCT	AAACATTTTA
560	570	580	590	600
CTATTACCCC	CTCTGGCAAA	ACTTCTTTTG	CAAAAGCCTC	TCGCTATTTT
610	620	630	640	650
GGTTTTATC	GTCGTCTGGT	AAACGAGGGT	TATGATAGTG	TTGCTCTTAC
660	670	680	690	700
TATGCCTCGT	AATTCCTTTT	GGCGTTATGT	ATCTGCATTA	GTTGAATGTG
710	720	730	740	750
GTATTCCTAA	ATCTCAACTG	ATGAATCTTT	CTACCTGTAA	TAATGTTGTT
760	770	780	790	800
CCGTTAGTTC	GTTTTATTAA	CGTAGATTTT	TCTTCCCAAC	GTCCTGACTG
810	820	830	840	850
GTATAATGAG	CCAGTTCTTA	AAATCGCATA	AGGTAATTCA	CAATGATTAA
860	870	880	890	900

FIGURE 9 (continued)

AGTTGAAATT	AAACCATCTC	AAGCCCAATT	TACTACTCGT	TCCTGGTGGT
910	920	930	940	950
CTCGTCAGGG	CAAGCCCTAT	TCACTGAATG	AGCAGCCTTG	TTACGTTGAT
960	970	980	990	1000
TTGGGTAATG	AATATCCGGT	TCTTGTCAAG	ATTACTCTTG	ATGAAGGTCA
1010	1020	1030	1040	1050
GCCAGCCTAT	GCGCCTGGTC	TGTACACCCT	TCATCTGTCC	TCCTTCAAAG
1060	1070	1080	1090	1100
TTGGTCAGTT	CGGTTCCCTT	ATGATTGACC	GTCTGCCGCT	CGTTCCGGCT
1110	1120	1130	1140	1150
AAGTAACATG	GAGCAGGTCT	CGGATTTCGA	CACAATTTAT	CAGGCGATGA
1160	1170	1180	1190	1200
TACAAATCTC	CGTTGTACTT	TGTTTCSCGC	TTGGTATAAT	CGCTGGGGGT
1210	1220	1230	1240	1250
CAAAGATGAG	TGTTTTAGTG	TATTCTTTCG	CCTCTTTCGT	TTTAGGTTGG
1260	1270	1280	1290	1300
TGCCCTTCGTA	GTGGCATTAC	GTATTTTACC	CGTTTAATGC	AAACTTCCTC
1310	1320	1330	1340	1350
ATGAAAAAGT	CTTTAGTCTT	CAAAGCCTCT	GTAGCCCTTG	CTACCCCTCGT
1360	1370	1380	1390	1400
TCCGATGCTG	TCTTTCGCTG	CTGAGGGTGA	CGATCCCGCA	AAAGCGGCCT
1410	1420	1430	1440	1450
TTACTCCCT	GCAAGCCTCA	GCGACCSAAT	ATATCGGTTA	TGCGTGGGCG
1460	1470	1480	1490	1500
ATGGTTGTGG	TCATTGTCTG	CGCAACTATC	GGTATCAAGC	TGTTTAAGAA
1510	1520	1530	1540	1550
ATTCACCTCG	AAAGCAAGCT	GATAAACCSA	TACAATTAAA	GGCTCCCTTTT
1560	1570	1580	1590	1600
GGAGCCTTTT	TTTTTGGAGA	TTTTCAACGT	GAAAAAATTA	TTATTCGCAA
1610	1620	1630	1640	1650
TTCCTTTAGT	TGTTCCCTTC	TATTCTCACT	CCGCTGAATC	TGTTGAAAGT
1660	1670	1680	1690	1700
TGTTTAGCAA	AACCCCATAC	AGAAAATTCA	TTTACTTACG	TCTGGAAAGA
1710	1720	1730	1740	1750
CGACAAACT	TTAGATCGTT	ACGCTAACTA	TGAGGGTTGT	CTGTGGAAATG
1760	1770	1780	1790	1800
CTACAGGCGT	TGTAGTTTGT	ACTGGTGACG	AAACTCAGTG	TTACGGTACA
1810	1820	1830	1840	1850
TGGGTTCCCTA	TTGGGCTTGC	TATCCCTGAA	AATGAGGGTG	GTGGCTCTGA
1860	1870	1880	1890	1900
GGGTGGCGGT	TCTGAGGGTG	GCGGTTCTGA	GGGTGGCGGT	ACTAAACCTC
1910	1920	1930	1940	1950
CTGAGTACGG	TGATACACCT	ATTCCGGGCT	ATACTTATAT	CAACCCCTCTC
1960	1970	1980	1990	2000
GACGGCACTT	ATCCGCCTGG	TACTGAGCAA	AACCCCGCTA	ATCCTAATCC
2010	2020	2030	2040	2050
TTCTCTTGAG	GAGTCTCAGC	CTCTTAATAC	TTTCATGTTT	CAGAAATAAT
2060	2070	2080	2090	2100
GGTTCCGAAA	TAGGCAGGGG	GCATTAACTG	TTTATACGGG	CACTGTTACT
2110	2120	2130	2140	2150
CAAGGCACTG	ACCCCGTTAA	AACTTATTAC	CAGTACACTC	CTGTATCATC
2160	2170	2180	2190	2200
AAAAGCCATG	TATGACGCTT	ACTGGAACGG	TAAATTCAGA	GACTGCGGCTT
2210	2220	2230	2240	2250

FIGURE 5 (continued)

TCCATTCTGG	CTTTAATGAA	GATCCATTGG	TTTGTGAATA	TCAAGGCCAA
2260	2270	2280	2290	2300
TCGTCTGACC	TGCCTCAACC	TCCTGTCAAT	GCTGGCGGCG	GCTCTGGTGG
2310	2320	2330	2340	2350
TGGTTCTGGT	GGCGGCTCTG	AGGGTGGTGG	CTCTGAGGGT	GGCGGTTCTG
2360	2370	2380	2390	2400
AGGGTSGCGG	CTCTGAGGGA	GGCGGTTCCG	GTGGTGGCTC	TGGTTCCGGT
2410	2420	2430	2440	2450
GATTTTGATT	ATGAAAAGAT	GGCAAACGCT	AATAAGGGGG	CTATGACCGA
2460	2470	2480	2490	2500
AAATGCCGAT	GAAAACGCGC	TACAGTCTGA	CGCTAAAGGC	AAACTTGATT
2510	2520	2530	2540	2550
CTGTGCGTAC	TGATTACGGT	GCTGCTATCG	ATGGTTTCAT	TGGTGACGTT
2560	2570	2580	2590	2600
TCCGGCCTTG	CTAATGGTAA	TGGTGCTACT	GGTGATTTTG	CTGGCTCTAA
2610	2620	2630	2640	2650
TTCCCAAATG	GCTCAAGTCG	GTGACGGTGA	TAATTCACCT	TTAATGAATA
2660	2670	2680	2690	2700
ATTTCCGTCA	ATATTTACCT	TCCCTCCCTC	AATCGGTTGA	ATGTCGCCCT
2710	2720	2730	2740	2750
TTTGTCTTTA	GCGCTGGTAA	ACCATATGAA	TTTTCTATTG	ATTGTGACAA
2760	2770	2780	2790	2800
AATAAACTTA	TTCCGTGGTG	TCTTTGCGTT	TCTTTTATAT	GTTGCCACCT
2810	2820	2830	2840	2850
TTATGTATGT	ATTTTCTACG	TTTGCTAACA	TACTGCGTAA	TAAGGAGTCT
2860	2870	2880	2890	2900
TAATCATGCC	AGTTCTTTTG	GGTATTCCGT	TATTATTGCG	TTTCCCTCGGT
2910	2920	2930	2940	2950
TTCCCTCTGG	TAACTTTGTT	CGGCTATCTG	CTTACTTTTC	TTAAAAAGGG
2960	2970	2980	2990	3000
CTTCGGTAAG	ATAGCTATTG	CTATTTCAAT	GTTTCTTGCT	CTTATTATTG
3010	3020	3030	3040	3050
GGCTTAACTC	AATTCCTGTG	GGTTATCTCT	CTGATATTAG	CGCTCAATTA
3060	3070	3080	3090	3100
CCCTCTGACT	TTGTTCAAGG	TGTTCAAGTTA	ATTCTCCCGT	CTAATGCGCT
3110	3120	3130	3140	3150
TCCCTGTTTT	TATGTTATTC	TCTCTGTAAA	GGCTGCTATT	TTCAATTTTTG
3160	3170	3180	3190	3200
ACGTTAAACA	AAAAATCGTT	TCTTATTTGG	ATTGGGATAA	ATAATATGGC
3210	3220	3230	3240	3250
TGTTTATTTT	GTAAGTGGCA	AATTAGGCTC	TGGAAAGACG	CTCGTTAGCG
3260	3270	3280	3290	3300
TTGGTAAGAT	TCAGGATAAA	ATTGTAGCTG	GGTGCAAAAT	AGCAACTAAT
3310	3320	3330	3340	3350
CTTGATTTAA	GGCTTCAAAA	CCTCCCGCAA	GTCGGGAGGT	TCGCTAAAC
3360	3370	3380	3390	3400
GCCTCGCGTT	CTTAGAATAC	CGGATAAGCC	TTCTATATCT	GATTTGCTTG
3410	3420	3430	3440	3450
CTATTGGGCG	CGGTAAAGAT	TGCTACGATG	AAAAATAAAA	CGGCTTGCTT
3460	3470	3480	3490	3500
GTTCTCGATG	AGTGCAGTAC	TTGGTTTAAAT	ACCCGTTCTT	GGAATGATAA
3510	3520	3530	3540	3550
GGAAAGACAG	CCGATTATTS	ATGGGTTTCT	ACATGCTCGT	AAATTAGTAT
3560	3570	3580	3590	3600

FIGURE 9 (continued)

GGGATATTAT	TTTTCTGGT	CAGGACTTAT	CTATTGTTGA	TAAACAGGCG
3610	3620	3630	3640	3650
CGTTCTGCAT	TAGCTGAACA	TGTTGTTTAT	TGTCGTCGTC	TGGACAGAAT
3660	3670	3680	3690	3700
TACTTTACCT	TTTGTGCGTA	CTTTATATTC	TCTTATTACT	GGCTCGAAAA
3710	3720	3730	3740	3750
TGCCCTCTGCC	TAAATTACAT	GTTGGCGTTG	TTAAATATGG	CGATTCTCAA
3760	3770	3780	3790	3800
TTAAGCCCTA	CTGTTGAGCG	TTGGCTTTAT	ACTGGTAAGA	ATTTGTATAA
3810	3820	3830	3840	3850
CGCATATGAT	ACTAAACAGG	CTTTTCTAG	TAATTATGAT	TCCGGTGTCT
3860	3870	3880	3890	3900
ATTCTTATTT	AACGCCTTAT	TTATCACACG	GTCGGTATTT	CAAACCATTA
3910	3920	3930	3940	3950
AATTTAGGTC	AGAAGATGAA	ATTAACATAA	ATATATTTGA	AAAAGTTTTT
3960	3970	3980	3990	4000
TCGCGTTCTT	TGCTCTCCGA	TTGGATTGTC	ATCAGCATTT	ACATATAGTT
4010	4020	4030	4040	4050
ATATAACCCA	ACCTAAGCCG	GAGGTTAAAA	AGGTAGTCTC	TCAGACCTAT
4060	4070	4080	4090	4100
GATTTTGATA	AATTCACAT	TGACTCTTCT	CAGCGTCTTA	ATCTAAGCTA
4110	4120	4130	4140	4150
TCGCTATGTT	TTCAAGGATT	CTAAGGGAAA	ATTAATTAAT	AGCGACGATT
4160	4170	4180	4190	4200
TACAGAAGCA	AGGTTATTCA	CTCACATATA	TTGATTTATG	TACTGTTTCC
4210	4220	4230	4240	4250
ATTAAAAAAG	GTAATTCAAA	TGAAATTGTT	AAATGTAATT	AATTTTGTTT
4260	4270	4280	4290	4300
TCTTGATGTT	TGTTTCATCA	TCTTCTTTTG	CTCAGGTART	TGAAATGAAT
4310	4320	4330	4340	4350
AATTCGCCTC	TGCGCGATTT	TGTAACCTGG	TATTCAAAGC	AATCAGGCGA
4360	4370	4380	4390	4400
ATCCGTTATT	GTTTCTCCCG	ATGTAAAAGG	TACTGTTACT	GTATATTCAAT
4410	4420	4430	4440	4450
CTGACGTTAA	ACTTGAAAAA	CTACGCAATT	TCITTATTTT	TGTTTTACGT
4460	4470	4480	4490	4500
GCTAATAATT	TTGATATGGT	TGGTTCAATT	CCTTCCATAA	TTCAGAAGTA
4510	4520	4530	4540	4550
TAATCCAAAC	AATCAGGTAT	ATATTGATGA	ATTGCCATCA	TCTGATAATC
4560	4570	4580	4590	4600
AGGAAATGTA	TGATAATTCC	GCTCCTTCTG	GTGGTTTCTT	TGTTCCGCAA
4610	4620	4630	4640	4650
AATGATAATG	TACTCAAAAC	TTTTAAAAAT	AATAACGTTT	GGGCAAAGGA
4660	4670	4680	4690	4700
TTTAAATCGA	GTTGTGGAAT	TGTTTGTAAA	GTCTAATACT	TCTAAATCCT
4710	4720	4730	4740	4750
CAATATGATT	ATCTATTGAC	GGCTCTAATC	TATTAGTTGT	TAGTGCACCT
4760	4770	4780	4790	4800
AAAGATATTT	TAGATAACCT	TCCTCAATTC	CTTTCTACTG	TTGATTTGCC
4810	4820	4830	4840	4850
AACGTACCAG	ATATTGATTG	AGGGTTTGGT	ATTTGAGGTT	CAGCAAGGTC
4860	4870	4880	4890	4900
ATGCTTTAGA	TTTTTTCATG	GCTGCTGGCT	CTCAGCGTGG	CACTGTTGCA
4910	4920	4930	4940	4950

FIGURE 9 (continued)

GGCGGTGTTA	ATACTGACCG	CCTCACCTCT	GTTTTATCTT	CTGCTGGTGG
4960	4970	4980	4990	5000
TTCGTTCCGG	ATTTTAAATG	GCGATGTTTT	AGGGCTATCA	GTTCGCGCAT
5010	5020	5030	5040	5050
TAAAGACTAA	TAGCCATTCA	AAAATATTGT	CTGTGCCACG	TATTCTTACG
5060	5070	5080	5090	5100
CITTCAGGTC	AGAAGGGTTC	TATCTCTGTT	GGCCAGAATG	TCCCTTTTAT
5110	5120	5130	5140	5150
TACTGGTCGT	GTGACTGGTG	AATCTGCCAA	TGTAAATAAT	CCATTTCAGA
5160	5170	5180	5190	5200
CGATTGAGCG	TCAAAATGTA	GGTATTTCCA	TGAGCGTTTT	TCCTGTTGCA
5210	5220	5230	5240	5250
ATGGCTGGCG	GTAATATTGT	TCTGGATATT	ACCAGCAAGG	CCGATAGTTT
5260	5270	5280	5290	5300
GAGTTCCTCT	ACTCAGGCAA	GTGATGTTAT	TACTAATCAA	AGAAGTATTG
5310	5320	5330	5340	5350
CTACAACGGT	TAATTTGCGT	GATGGACAGA	CTCTTTTACT	CGGTGGCCTC
5360	5370	5380	5390	5400
ACTGATTATA	AAAACACTTC	TCAAGATTCT	GGCGTACCGT	TCCTGTCTAA
5410	5420	5430	5440	5450
AATCCCTTTA	ATCGGCCTCC	TGTTTAGCTC	CCGCTCTGAT	TCCAACGAGG
5460	5470	5480	5490	5500
AAAGCACGTT	ATACGTGCTC	GTCAAAGCAA	CCATAGTACG	CGCCCTGTAG
5510	5520	5530	5540	5550
CGGCGCATT	AGCGCGGCGG	GTGTGGTGGT	TACGCGCAGC	GTGACCGCTA
5560	5570	5580	5590	5600
CACTTGCCAG	CGCCCTAGCG	CCCGCTCCTT	TCGCTTCTCT	CCCTTCCTTT
5610	5620	5630	5640	5650
CTCGCCACGT	TCGCCGGCTT	TCCCCGTCAA	GCTCTAAATC	GGGGGCTCCC
5660	5670	5680	5690	5700
TTTAGGGTTC	CGATTTAGTG	CTTTACGGCA	CCTCGACCCC	AAAAAAGTTG
5710	5720	5730	5740	5750
ATTTGGGTGA	TGGTTCACGT	AGTGGGCCAT	CGCCCTGATA	GACGGTTTTT
5760	5770	5780	5790	5800
CGCCCTTTGA	CGTTGGAGTC	CACGTTCTTT	AATAGTGGAC	TCTTGTTCCA
5810	5820	5830	5840	5850
AACTGGAACA	ACACTCAACC	CTATCTCGGG	CTATTCTTTT	GATTTATAAG
5860	5870	5880	5890	5900
GGATTTTGCC	GATTTTCGGAA	CCACCATCAA	ACAGGATTTT	CGCCTGCTGG
5910	5920	5930	5940	5950
GGCAAACCAG	CGTGGACCGC	TTGCTGCAAC	TCTCTCAGGG	CCAGGCGGTG
5960	5970	5980	5990	6000
AAGGGCAATC	AGCTGTTGCC	CGTCTCGCTG	GTGAAAAGAA	AAACCACCCCT
6010	6020	6030	6040	6050
GGCGCCCAAT	ACGCAAACCG	CCTCTCCCCG	CGCGTTGGCC	GATTCATTAA
6060	6070	6080	6090	6100
TCCAGCTGGC	ACGACAGGTT	TCCCGACTGG	AAAGCGGGCA	GTGAGCGCAA
6110	6120	6130	6140	6150
CGCAATTAAT	GTGAGTTACC	TCACTCATTA	GGCACCCGAG	GCTTTTACACT
6160	6170	6180	6190	6200
TTATGCTTCC	GGCTCGTATG	TTGTGTGGAA	TTGTGAGCGG	ATAACAATTT
6210	6220	6230	6240	6250
CACACAGGAA	ACAGCTATGA	CCATGATTAC	GAATTCGAGC	TCGCCCCGGG
6260	6270	6280	6290	6300

FIGURE 9 (continued)

ATCTGCCTGA	ATAGGTACGA	TTTACTAACT	GGAAGAGGCA	CTAAATGAAC
6310	6320	6330	6340	6350
ACGATTAACA	TCGCTAAGAA	CGACTTCTCT	GACATCGAAC	TGGCTGCTAT
6360	6370	6380	6390	6400
CCCGTTCAAC	ACTCTGGCTG	ACCATTACGG	TGAGCGTTTA	GCTCGCGAAC
6410	6420	6430	6440	6450
AGTTGGCCCT	TGAGCATGAG	TCTTACGAGA	TGGGTGAAGC	ACGCTTCCGC
6460	6470	6480	6490	6500
AAGATGTTTG	AGCGTCAACT	TAAAGCTGGT	GAGGTTGCGG	ATAACGCTGC
6510	6520	6530	6540	6550
CGCCAAGCCT	CTCATCACTA	CCCTACTCCC	TAAGATGATT	GCACGCATCA
6560	6570	6580	6590	6600
ACGACTGGTT	TGAGGAAGTG	AAAGCTAAGC	GCGGCAAGCG	CCCGACAGCC
6610	6620	6630	6640	6650
TTCCAGTTCC	TGCAAGAAAT	CAAGCCGGAA	GCCGTAGCGT	ACATCACCAT
6660	6670	6680	6690	6700
TAAGACCACT	CTGGCTTGCC	TAACCAGTGC	TGACAAATCA	ACCGTTCAGG
6710	6720	6730	6740	6750
CTGTAGCAAG	CGCAATCGGT	CGGGCCATTG	AGGACGAGGC	TCGCTTCGGT
6760	6770	6780	6790	6800
CGTATCCGTG	ACCTTGAAGC	TAAGCACTTC	AAGAAAAACG	TTGAGGAACA
6810	6820	6830	6840	6850
ACTCAACAAG	CGCGTAGGGC	ACGTCTACAA	GAAAGCATTG	ATGCAAGTTG
6860	6870	6880	6890	6900
TCGAGGCTGA	CATGCTCTCT	AAGGGTCTAC	TCGGTGGCGA	GGCGTGGTCT
6910	6920	6930	6940	6950
TCGTGGCATA	AGGAAGACTC	TATTCATGTA	GGAGTACGCT	GCATCGAGAT
6960	6970	6980	6990	7000
GCTCATTGAG	TCAACCGGAA	TGGTTAGCTT	ACACCGCCAA	AATGCTGGCG
7010	7020	7030	7040	7050
TAGTAGGTCA	AGACTCTGAG	ACTATCGAAC	TCGCACCTGA	ATACGCTGAG
7060	7070	7080	7090	7100
GCTATCGCAA	CCCGTGCAGG	TGCGCTGGCT	GGCATCTCTC	CGATGTTCCA
7110	7120	7130	7140	7150
ACCTTGCCTA	GTTCTCTCCT	AGCCGTGGAC	TGGCATTACT	GGTGGTGGCT
7160	7170	7180	7190	7200
ATTGGGCTAA	CGGTCTGCTG	CCTCTGGCGC	TGGTGCCTAC	TCACAGTAAG
7210	7220	7230	7240	7250
AAAGCACTGA	TGCGCTACGA	AGACGTTTAC	ATGCCTGAGG	TGTACAAAGC
7260	7270	7280	7290	7300
GATTAACATT	GCGCAAAACA	CCGCATGGAA	AATCAACAG	AAAGTCCTAG
7310	7320	7330	7340	7350
CGGTGCGCAA	CGTAATCACC	AAGTGGAGC	ATTGTCCGGT	CGAGGACATC
7360	7370	7380	7390	7400
CCTGCGATTG	AGCGTGAAGA	ACTCCCGATG	AAACCGGAAG	ACATCGACAT
7410	7420	7430	7440	7450
GAATCCTGAG	GCTCTCACCG	CGTGGAAACG	TGCTGCCGCT	GCTGTGTACC
7460	7470	7480	7490	7500
GCAAGGACAA	GGCTCGCAAG	TCTCGCCGTA	TCAGCCTTGA	GTTTCATGCTT
7510	7520	7530	7540	7550
GAGCAAGCCA	ATAAGTTTGC	TAACCATAAG	GCCATCTGGT	TCCCTTACAA
7560	7570	7580	7590	7600
CATGGACTGG	CGCGGTCTGT	TTTACGCTGT	GTCATGTGTC	AACCCGCAAG
7610	7620	7630	7640	7650

FIGURE 9 (continued)

GTAACGATAT	GACCAAAGSA	CTGCTTACGC	TGGCGAAAGG	TAAACCAATC
7650	7670	7680	7690	7700
GGTAAGGAAG	GTTACTACTG	GCTGAARAATC	CACGGTGCAA	ACTGTGCGGG
7710	7720	7730	7740	7750
TGTCGATAAG	GTTCCGTTCC	CTGAGCGCAT	CAAGTTCATT	GAGGAAAACC
7760	7770	7780	7790	7800
ACGGAACAT	CATGGCTTGC	GCTAAGTCTC	CACTGGAGAA	CACTTGGTGG
7810	7820	7830	7840	7850
GCTGAGCAAG	ATTCTCCGTT	CTGCTTCCTT	GCGTTCGTCT	TTGAGTACGC
7860	7870	7880	7890	7900
TGGGGTACAG	CACCACGGCC	TGAGCTATAA	CTGCTCCCTT	CCGCTGGCGT
7910	7920	7930	7940	7950
TTGACGGGTC	TTGCTCTGGC	ATCCAGCACT	TCTCCGCGAT	GCTCCGAGAT
7960	7970	7980	7990	8000
GAGGTAGGTG	GTCGCGCGGT	TAAGTTGCTT	CCTAGTGAAA	CCGTTTCAGGA
8010	8020	8030	8040	8050
CATCTACGGG	ATTGTTGCTA	AGAAAGTCAA	CGAGATTCTA	CAAGCAGACG
8060	8070	8080	8090	8100
CAATCAATGG	GACCGATAAC	GAAGTAGTTA	CCGTGACCGA	TGAGAACACT
8110	8120	8130	8140	8150
GGTGAAATCT	CTGAGAAAGT	CAAGCTGGGC	ACTAAGGCAC	TGGCTGGTCA
8160	8170	8180	8190	8200
ATGGCTGGCT	TACGGTGTTA	CTCGCAGTGT	GACTAAGCGT	TCAGTCATGA
8210	8220	8230	8240	8250
CGCTGGCTTA	CGGGTCCAAA	GAGTTCGGCT	TCCGTCAACA	AGTGCTGGAA
8260	8270	8280	8290	8300
GATACCATTC	AGCCAGCTAT	TGATTCCGGC	AAGGGTCTGA	TGTTCACTCA
8310	8320	8330	8340	8350
GCCGAATCAG	GCTGCTGGAT	ACATGGCTAA	GCTGATTTGG	GAATCTGTGA
8360	8370	8380	8390	8400
GCGTGACGGT	GGTAGCTGCG	GTTGAAGCAA	TGAACTGGCT	TAAGTCTGCT
8410	8420	8430	8440	8450
GCTAAGCTGC	TGGCTGCTGA	GGTCAAAGAT	AAGAAGACTG	GAGAGATTCT
8460	8470	8480	8490	8500
TCGCAAGCGT	TGCGCTGTGC	ATTGGGTAAC	TCCTGATGGT	TTCCCTGTGT
8510	8520	8530	8540	8550
GGCAGGAATA	CAAGAAGCCT	ATTCAGACGC	GCTTGAACCT	GATGTTCCCTC
8560	8570	8580	8590	8600
GGTCAGTTCC	GCTTACAGCC	TACCATTAAC	ACCAACAAAG	ATAGCGAGAT
8610	8620	8630	8640	8650
TGATGCACAC	AAACAGGAGT	CTGGTATCJC	TCCTAACTTT	GTACACAGCC
8660	8670	8680	8690	8700
AAGACGGTAG	CCACCTTCGT	AAGACTGTAG	TGTGGGCACA	CGAGAAGTAC
8710	8720	8730	8740	8750
GGAATCGAAT	CTTTTGCACT	GATTCACGAC	TCCTTCGGTA	CCATTCCGGC
8760	8770	8780	8790	8800
TGACGCTGCG	AACCTGTTCA	AAGCAGTGCG	CGAAACTATG	GTTGACACAT
8810	8820	8830	8840	8850
ATGAGTCTTG	TGATGTACTG	GCTGATTTCT	ACGACCAGTT	CGCTGACCAG
8860	8870	8880	8890	8900
TTGCACGAGT	CTCAATTGGA	CAAAATGCCA	GCACTTCCGG	CTAAGGTTAA
8910	8920	8930	8940	8950
CTTGAACCTC	CGTGACATCT	TAGAGTCGGA	CTTCGCGTTC	GCGTAACGCC
8960	8970	8980	8990	9000

FIGURE 9 (continued)

AAATCAATAC	GACCCGGATC	GGTCGACCTG	CAGCCCAAGC	TTGGCACTGG
9010	9020	9030	9040	9050
CCGTCGTTTT	ACAACGTCGT	GACTGGGAAA	ACCCTGGCGT	TACCCAACTT
9060	9070	9080	9090	9100
AATCGCCTTG	CAGCACATCC	CCCCTTCGCC	AGCTGGCGTA	ATAGCGAAGA
9110	9120	9130	9140	9150
GGCCCCGACC	GATCGCCCTT	CCCAACAGTT	GCGTAGCCTG	AATGGCGAAT
9160	9170	9180	9190	9200
GGCGCTTTGC	CTGGTTTCCG	GCACCAGAAG	CGGTGCCGGA	AAGCTGGCTG
9210	9220	9230	9240	9250
GAGTGCATC	TTCTGAGGC	CGAQAACNGT	GTCGTCCCCT	CAAACCTGGCA
9260	9270	9280	9290	9300
GATGCACGGT	TACGATGCGC	CCATCTACAC	CAACGTAACC	TATCCCATTA
9310	9320	9330	9340	9350
CGGTCAATCC	GCCGTTTGTT	CCCACGGAGA	ATCCGACGGG	TTGTTACTCG
9360	9370	9380	9390	9400
CTCACATTTA	ATGTTGATGA	AAGCTGGCTA	CAGGAAGGCC	AGACGCGAAT
9410	9420	9430	9440	9450
TATTTTTGAT	GGCGTTCCTA	TTGGTTAAAA	AATGAGCTGA	TTTAACAAAA
9460	9470	9480	9490	9500
ATTTAACGCG	AATTTTAACA	AAATATTAAC	GTTTACAATT	TAAATATTTG
9510	9520	9530	9540	9550
CTTATACAAT	CTTCCTGTTT	TTGGGGCTTT	TCTGATTATC	AACCGGGGTA
9560	9570	9580	9590	9600
CATATGATTG	ACATGCTAGT	TTTACGATT	CCGTTTCATCG	ATTCTCTTGT
9610	9620	9630	9640	9650
TTGCTCCAGA	CTCTCAGGCA	ATGACCTGAT	AGCCTTTGTA	GATCTCTCAA
9660	9670	9680	9690	9700
AAATAGCTAC	CCTCTCCGGC	ATGAATTTAT	CAGCTAGAAC	GGTTGAATAT
9710	9720	9730	9740	9750
CATATTGATG	GTGATTGAC	TGTCTCCGGC	CTTTCTCACC	CTTTTGAATC
9760	9770	9780	9790	9800
TTTACCTACA	CATTACTCAG	GCATTGCATT	TAAAATATAT	GAGGGTTCTA
9810	9820	9830	9840	9850
AAAATTTTTA	TCCTTGCGTT	GAAATAAAGG	CTTCTCCCGC	AAAAGTATTA
9860	9870	9880	9890	9900
CAGGGTCATA	ATGTTTTTGG	TACAACCGAT	TTAGCTTTAT	GCTCTGAGGC
9910	9920	9930	9940	9950
TTTATTGCTT	AATTTTGCTA	ATTCTTTGCC	TTGCCTGTAT	GATTATTTGG

ATGTT

FIGURE 10

